

UNIVERSITÉ DE MONTRÉAL

CODAGE STÉRÉOSCOPIQUE AVEC RÉGION
D'INTÉRÊT

FRANÇOYS LABONTÉ

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE ET DE GÉNIE INFORMATIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE ÉLECTRIQUE)

FÉVRIER 1997



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-33009-5

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

CODAGE STÉRÉOSCOPIQUE AVEC RÉGION
D'INTÉRÊT

présentée par: LABONTÉ François

en vue de l'obtention du diplôme de: Philosophiæ Doctor

a été dûment acceptée par le jury d'examen constitué de:

M. SAWAN Mohamad, Ph.D., président

M. COHEN Paul, Ph.D., membre et directeur de recherche

M. FAUBERT Jocelyn, Ph. D., membre et codirecteur de recherche

M. DUBOIS Éric, Ph.D., membre

M. HERVÉ Jean-Yves, Ph. D., membre

Remerciements

Je tiens tout d'abord à remercier mon directeur de recherche, Paul Cohen, ainsi que mes deux codirecteurs de recherche, Chon Tam LeDinh et Jocelyn Faubert, pour le temps et les efforts qu'ils ont consacrés à ce projet. J'aimerais aussi remercier les membres du groupe de recherche en perception et robotique, plus spécialement Frédéric Labrosse, Christian Zanardi et Jean-Yves Hervé, pour leur support, leurs conseils et leur aide. Je tiens aussi à remercier les étudiants de l'Ecole d'Optométrie qui ont collaboré à la réalisation de l'étude psychophysique du chapitre 4 : Anne Bellefeuille, Lynda Bilodeau et Vasile Diaconu, ainsi que Magdalena Balazinska qui a accompli une partie de cette étude dans le cadre d'un stage d'été. Je voudrais également remercier mes parents, Denise et René, ma soeur, Anne-Marie, ainsi que mon amie Caroline pour leur encouragements soutenus au cours de ces dernières années. Finalement, je tiens aussi à remercier le FCAR et le réseau de centres d'excellence IRIS pour leur support financier.

Résumé

Nous proposons une nouvelle approche de compression de séquences stéréoscopiques qui compacte l'information stéréoscopique dans l'espace spectral d'un canal vidéo monoculaire et qui exploite la présence d'une région de fixation pour réduire la quantité d'information à transmettre. Puisque les acuités spatiales et stéréoscopiques du système visuel humain sont plus élevées autour du point de fixation des yeux et diminuent rapidement avec une augmentation de l'excentricité rétinienne, l'information de haute-fréquences n'est pas transmise pour tout le champ visuel, mais seulement à l'intérieur d'un sous-espace contenant la région de fixation. Cette approche de compression soulève deux problèmes fondamentaux : l'estimation d'une région de fixation et la modulation de la fidélité du codage en fonction de cette région. Ces deux problèmes sont abordés dans ce travail.

Dans le but de développer une stratégie d'estimation de région de fixation, nous avons accompli une étude psychophysique portant sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur, simulée à l'aide de stéréogrammes dynamiques. Cette étude portait sur deux as-

pects de la performance des sujets lors de l'accomplissement de la tâche choisie : les mouvements de leurs yeux en fonction de la présence de distracteurs et l'effet des conditions particulières associées au visionnement stéréoscopique. Les résultats de l'étude montrent que les fixations des sujets se trouvent majoritairement dans la région centrale de l'écran correspondant à la tâche principale ainsi que dans une zone située à droite de celle-ci, ce qui indique que la tâche étudiée peut être accomplie sans déplacement significatif de la région de fixation. Une explication reliée à la géométrie des systèmes de visualisation stéréoscopique est proposée pour tenir compte des fixations se trouvant à droite de la région centrale. En fonction des résultats de cette étude, une stratégie d'estimation de région de fixation est proposée. Elle consiste à favoriser tout d'abord la région associée à la tâche principale, puis ensuite une zone adjacente à cette région, située du côté opposé à celui de l'œil qui la regarde pour chaque canal stéréoscopique. Cette stratégie est incorporée aux méthodes de codage qui sont développées dans ce travail.

L'approche de codage qui est proposée consiste à décomposer les champs gauche et droit de chaque paire stéréoscopique en composantes de luminance passe-bas et passe-haut. Les composantes passe-haut sont par la suite limitées à une région de fixation, permettant ainsi une réduction de l'espace spectral qu'elles occupent. À l'aide de manipulations spectrales, les composantes de luminance sont combinées à celles de chrominance afin de former un signal vidéo composite dont l'occupation spectrale n'excède pas celle d'un signal vidéo monoculaire. Les *trous spectraux de Fukinuki*

sont exploités afin de positionner les composantes de chrominance des deux canaux stéréoscopiques dans différents quadrants du plan fréquentiel temporel-vertical. L'approche proposée est compatible avec le standard NTSC dans la mesure où la même sous-porteuse de chrominance et la même région spectrale pour les composantes de chrominance sont utilisées. Il est donc possible d'utiliser des décodeurs NTSC standards pour séparer les composantes de luminance de celles de chrominance. Un traitement additionnel doit par la suite être accompli afin de récupérer l'information correspondant aux canaux gauche et droit. Contrairement à la majorité des méthodes de codage stéréoscopiques présentées dans la littérature, l'approche proposée n'implique pas d'appariement entre les canaux stéréoscopiques.

Trois méthodes de codage, toutes basées sur les principes qui viennent d'être mentionnés, sont proposées afin d'exploiter pleinement la bande passante de différents systèmes de communication : (1) la méthode *studio* pour laquelle la limite spectrale horizontale correspond à la moitié de la fréquence d'échantillonnage des images originales, soit 7,16 MHz, (2) la méthode *diffusion* pour laquelle la limite spectrale horizontale correspond à 4,2 MHz, et (3) la méthode *intégrée* qui peut satisfaire simultanément les contraintes de bande passante des deux premières méthodes. Le format *studio* est approprié pour une utilisation, entre autres, avec des enregistreuses numériques D2 pour la vidéo composite, alors que le format *diffusion* s'applique à la transmission d'images.

Le niveau de qualité des images codées par les méthodes proposées a été évalué

de façon subjective selon un critère de visibilité d'artefacts par quelques sujets et de façon objective à l'aide d'une mesure du rapport signal à bruit. Les résultats sont satisfaisants et démontrent une meilleure qualité des images dans la région de fixation, tel que désiré. Toutes les méthodes de codage ont été élaborées avec des filtres numériques unidimensionnels séparables à réponse impulsionnelle finie dont le nombre de coefficients varie de 11 à 41, ce qui les rend appropriées pour une implantation matérielle efficace.

Les principales contributions de ce travail proviennent de (1) l'exploitation d'une région de fixation pour la compression spectrale de séquences stéréoscopiques, (2) la réalisation d'une étude psychophysique sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur, et (3) l'intégration des résultats de cette étude aux méthodes de codage développées. De plus, l'approche proposée se distingue par une compatibilité avec le standard NTSC et par son efficacité du fait qu'elle implique des opérations de filtrage simples et qu'elle ne nécessite pas d'appariement entre les canaux stéréoscopiques.

Abstract

We propose a new compression approach for interlaced stereoscopic sequences which compacts the stereoscopic information into the spectral space of monocular video channel, and which exploits the presence of a fixation region to reduce the amount of transmitted information. Since spatial and stereoscopic acuities of the human visual system are highest around the fixation point of the eyes and decrease rapidly with increasing retinal eccentricity, high-frequency information is not transmitted over the entire visual field, but only inside a subspace containing the fixation region. This compression approach raises two fundamental problems: the estimation of the fixation region and the modulation of the coding fidelity as a function of this region. These two problems are addressed in this work.

In order to develop a fixation region estimation strategy, we performed a psychophysical experiment on the visual strategies used during the accomplishment of a dynamic depth discrimination task, simulated with dynamic stereograms. This study addressed two aspects of the subjects performance during the accomplishment of the chosen task: their eye movements as a function of the presence of distractors, and

the effect of the specific conditions of stereoscopic viewing. Results of the study show that subjects fixations are mainly located in the central region of the screen corresponding to the main task, and inside a region located to the right of the central region. This indicates that the studied task can be accomplished without any significant displacement of the fixation region. An explanation related to the geometry of stereoscopic visualization systems is proposed to take into account the fixations located to the right of the central region. A fixation region estimation strategy which takes into accounts the findings of this study is proposed. It consists of favoring first the region associated with the main task, and next a region located on the opposite side of the viewing eye of each stereoscopic channel. This strategy is incorporated into the coding methods developed in this work.

The proposed coding approach consists of decomposing the left and right fields of each stereoscopic pair into lowpass and highpass luminance components. The highpass components are then limited to a fixation region, thus allowing a reduction of the spectral space they occupy. Luminance components are then combined with chrominance components by means of spectral manipulations to form a composite video signal whose spectral occupation does not exceed the one of a monocular video channel. The *spectral Fukinuki holes* are exploited to position the chrominance components of the two stereoscopic channels into different quadrants in the temporal-vertical frequency plane. The proposed approach is compatible with the NTSC standard in the sense that the same chrominance subcarrier and the same chrominance spectral

region are used. It is therefore possible to use standard NTSC decoders to separate luminance from chrominance. Additional processing must then be performed to obtain the information corresponding to the left and right channels. Contrary to most stereoscopic coding methods presented in the literature, the proposed approach does not involve a matching between the stereoscopic channels.

Three coding methods, all based on the above-mentioned principles, are proposed in order to fully exploit the bandwidth of different communication systems: (1) the *studio* method for which the horizontal spectral limit corresponds to half the sampling frequency of the original images, that is 7.16 MHz, (2) the *broadcast* method for which the horizontal spectral limit is fixed to 4.2 MHz, and (3) the *integrated* method which can simultaneously satisfy the bandwidth constraint of the two first approaches. The *studio* format is appropriate to be used with, among others, D2 numerical recorders for composite video signals, while the *broadcast* format is suitable for image transmission.

The quality level of the images coded with the proposed methods has been evaluated in a subjective way according to an impairment visibility criterion by a few subjects, and in an objective way according to a signal-to-noise ratio measure. Results are satisfying and show a better quality level in the fixation region, as desired. All coding methods were implemented with 1D separable finite impulse response numerical filters with a number of coefficients ranging from 11 to 41, which makes them appropriate for an efficient hardware implementation.

The main contributions of this work come from: (1) the exploitation of a fixation

region for the spectral compression of stereoscopic sequences, (2) the realization of a psychophysical study on visual strategies used during the accomplishment of a dynamic depth discrimination task, and (3) the integration of the results of this study into the coding method which have been developed. Moreover, the proposed approach stands out by its compatibility with the NTSC standard and by its efficiency due to the fact that it implies simple filtering operations and does not necessitate a matching between the stereoscopic channels.

Table des matières

Remerciements	iv
Résumé	v
Abstract	ix
Table des matières	xiii
Liste des tableaux	xvii
Liste des figures	xxi
Liste des annexes	xxx
Introduction	1
1 Un survol bibliographique des approches de codage stéréoscopique	10
1.1 Théorie de la suppression et codage asymétrique	14
1.2 Théorie de la fusion et codage symétrique	15

1.3	Exploitation simultanée du mouvement et de la stéréoscopie	16
2	L'exploitation d'une région de fixation en codage	21
3	Un aperçu général de l'approche de codage proposée	29
3.1	Contexte et justification	29
3.2	Système de transmission d'images	36
4	Les stratégies visuelles pour une tâche de discrimination de pro- fondeur	38
4.1	Description de l'étude	38
4.2	Résultats	42
4.2.1	Direction du mouvement des distracteurs	42
4.2.2	Type du mouvement des distracteurs	43
4.2.3	Distributions des fixations	44
4.2.4	Effet d'apprentissage	44
4.2.5	Calcul du pointage	46
4.3	Discussion	46
5	Les méthodes de codage	51
5.1	Estimation de la région de fixation	52
5.2	Méthode studio : luminance limitée à $2f_{sp} = 7,16$ MHz	55
5.2.1	Traitement des composantes de luminance	55
5.2.2	Traitement spatial des composantes de chrominance	62

	xv
5.2.3 Traitement spatio-temporel des composantes de chrominance .	62
5.2.4 Reconstruction des composantes	64
5.3 Méthode diffusion : luminance limitée à $f_l = 4,2$ MHz	72
5.3.1 Traitement des composantes de luminance	73
5.3.2 Traitement des composantes de chrominance	76
5.3.3 Reconstruction des composantes	81
5.4 Méthode intégrée : luminance limitée à $2f_{sp} = 7,16$ MHz ou $f_l = 4,2$ MHz	83
5.5 Résumé des 3 méthodes	90
6 Les résultats expérimentaux	96
6.1 Illustration des étapes de traitement	100
6.2 Evaluation subjective	111
6.3 Evaluation objective	120
6.3.1 Transmission des composantes de luminance passe-bas	120
6.3.2 Transmission des composantes de luminance de la région de fixation	124
6.3.3 Transmission des composantes de chrominance	134
6.4 Discussion	139
Conclusion	141
Bibliographie	146

Annexes	167
--------------------------	------------

Liste des tableaux

5.1	Caractéristiques du traitement de luminance avec les différentes méthodes de codage.	92
5.2	Caractéristiques du traitement de chrominance avec les différentes méthodes de codage.	93
6.1	Echelle de visibilité de dégradations	118
6.2	Appréciation moyenne des six méthodes de codage	119
6.3	RSBC entre les composantes de luminance originales et passe-bas avant la formation du signal vidéo composite.	122
6.4	RSBC entre les composantes de luminance passe-bas originales et passe-bas reconstruites.	123
6.5	RSBC entre les composantes de luminance originales et passe-bas reconstruites.	125
6.6	RSBC entre les composantes de luminance originales et passe-bas de la région de fixation avant la formation du signal vidéo composite. . .	127

6.7	RSBC entre les composantes de luminance originales et la somme des composantes passe-bas et passe-haut de la région de fixation avant la formation du signal vidéo composite.	128
6.8	RSBC entre les composantes de luminance passe-bas originales et passe-bas reconstruites de la région de fixation.	130
6.9	RSBC entre la somme des composantes de luminance passe-bas et passe-haut originales de la région de fixation et la somme des composantes passe-bas et passe-haut reconstruites de la région de fixation.	131
6.10	RSBC entre les composantes de luminance originales et passe-bas reconstruites de la région de fixation.	133
6.11	RSBC entre les composantes de luminance originales et reconstruites de la région de fixation.	134
6.12	RSBC entre les composantes de chrominance originales et passe-bas avant la formation du signal vidéo composite.	135
6.13	RSBC entre les composantes de chrominance passe-bas originales et passe-bas reconstruites.	137
6.14	RSBC entre les composantes de chrominance originales et passe-bas reconstruites.	138
B.1	Valeurs représentatives de limites de disparité pour la fusion binoculaire, la stéréopsie obligatoire et la stéréopsie qualitative	184

C.1	Pourcentage de fixations d'un sujet dans les zones de la Figure 4.1 pour les 8 directions des distracteurs.	197
C.2	Pourcentage de fixations dans les zones de la Figure 4.1 pour le premier essai avec les distracteurs possédant un mouvement fronto-parallèle. .	199
C.3	Pourcentage de fixations dans les zones de la Figure 4.1 pour le second essai avec les distracteurs possédant un mouvement en profondeur. . .	200
C.4	Pourcentage de fixations dans les zones de la Figure 4.1 pour le premier essai avec les distracteurs possédant un mouvement en profondeur. . .	201
C.5	Pourcentage de fixations dans les zones de la Figure 4.1 pour le second essai avec les distracteurs possédant un mouvement fronto-parallèle. .	202
C.6	Différence de pourcentage entre les données moyennes des tableaux C.2, C.3, C.4 et C.5	203
C.7	Pourcentage moyen de fixations dans les zones de la Figure 4.1 pour les deux principales stratégies visuelles.	205
C.8	Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la même stratégie visuelle lors de leurs deux essais.	208
D.1	Filtres passe-bas de luminance pour la subdivision en composantes passe-bas et passe-haut.	218
D.2	Filtres pour le traitement des composantes de luminance passe-haut avec les méthodes <i>diffusion</i> et <i>intégrée</i>	219

D.3	Filtres passe-bande pour le traitement de la première sous-composante de luminance passe-haut avec les méthodes <i>diffusion</i> et <i>intégrée</i>	220
D.4	Filtres passe-bande pour le traitement de la seconde sous-composante de luminance passe-haut avec les méthodes <i>diffusion</i> et <i>intégrée</i>	220
D.5	Filtres passe-bas et passe-haut pour le traitement des composantes de chrominance.	221
D.6	Filtres passe-bande pour le traitement des composantes de chrominance.	221
D.7	Filtres passe-bande pour le traitement des composantes de chrominance et pour la séparation de l'information de chrominance de celle de luminance.	222
D.8	Séquence des composantes de chrominance en fonction de la phase des sous-porteuses principale et auxiliaire.	223
E.1	Appréciation du niveau de qualité des images pour le sujet CZ. . . .	251
E.2	Appréciation du niveau de qualité des images pour le sujet JYH. . . .	252
E.3	Appréciation du niveau de qualité des images pour le sujet AB. . . .	252

Liste des figures

3.1	Occupation spectral d'un champ du signal NTSC dans le volume fréquentiel horizontal-vertical-temporel.	34
3.2	Occupation spectral d'un champ du signal NTSC dans les plans fréquentsiels horizontal-vertical et temporel-vertical.	35
3.3	Système de transmission d'images.	37
4.1	Subdivision de l'écran en 21 zones.	43
4.2	Distribution des pourcentages de fixation dans les zones de la figure 4.1 pour les deux principales stratégies visuelles.	45
4.3	Géométrie des systèmes de visualisation stéréoscopique.	49
5.1	Division de l'espace spectral avec la méthode <i>studio</i>	56
5.2	Décomposition de l'information de luminance en composantes passe-bas et passe-haut avec la méthode <i>studio</i>	59
5.3	Formation du signal vidéo composite avec la méthode <i>studio</i>	60

5.4	Traitement des composantes passe-haut dans le domaine spectral avec la méthode <i>studio</i>	61
5.5	Traitement des composantes de chrominance avec la méthode <i>studio</i>	63
5.6	Occupation spectrale des composantes de chrominance avec la méthode <i>studio</i>	63
5.7	Traitement des composantes de chrominance avec la méthode <i>studio</i> avec trous de Fukinuki.	65
5.8	Formation du signal vidéo composite avec la méthode <i>studio</i> avec trous de Fukinuki.	65
5.9	Occupation spectrale des composantes de chrominance avec la méthode <i>studio</i> avec trous de Fukinuki.	66
5.10	Séparation des composantes de luminance et de de chrominance à partir du signal vidéo composite.	67
5.11	Séparation des composantes de luminance avec la méthode <i>studio</i>	68
5.12	Reconstruction des composantes de luminance passe-haut avec la méthode <i>studio</i>	69
5.13	Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode <i>studio</i>	70
5.14	Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode <i>studio</i> avec trous de Fukinuki.	71

5.15	Reconstruction des composantes de chrominance avec la méthode <i>studio</i> .	72
5.16	Division de l'espace spectral avec la méthode <i>diffusion</i>	74
5.17	Décomposition de l'information de luminance en composantes passe- bas et passe-haut avec la méthode <i>diffusion</i>	75
5.18	Traitement des composantes passe-haut du canal gauche avec la mé- thode <i>diffusion</i>	75
5.19	Traitement des composantes passe-haut du canal droit avec la méthode <i>diffusion</i>	76
5.20	Formation du signal vidéo composite avec la méthode <i>diffusion</i>	77
5.21	Traitement des composantes passe-haut dans le domaine spectral avec la méthode <i>diffusion</i>	78
5.22	Formation du signal vidéo composite avec la méthode <i>diffusion</i> avec trous de Fukinuki.	79
5.23	Occupation spectrale des composantes de chrominance avec la méthode <i>diffusion</i> avec trous de Fukinuki.	80
5.24	Traitement des composantes de chrominance du canal gauche avec la méthode <i>diffusion</i>	81
5.25	Traitement des composantes de chrominance du canal droit avec la méthode <i>diffusion</i>	82
5.26	Occupation spectrale des composantes de chrominance avec la méthode <i>diffusion</i>	82

5.27	Division de la région de fixation en 2 sous-régions	84
5.28	Division de l'espace spectral avec la méthode <i>intégrée</i>	86
5.29	Formation du signal vidéo composite avec la méthode <i>intégrée</i>	87
5.30	Décomposition de l'information de luminance en composantes passe- bas et passe-haut avec la méthode <i>intégrée</i>	88
5.31	Traitement des composantes passe-haut de la sous-région adjacente du canal gauche avec la méthode <i>intégrée</i>	89
5.32	Traitement de la composante passe-bas et des composantes passe-haut de la sous-région centrale du canal droit avec la méthode <i>intégrée</i> . . .	89
5.33	Traitement des composantes passe-haut de la sous-région adjacente du canal droit avec la méthode <i>intégrée</i>	90
5.34	Principales étapes de traitement avec la méthode <i>studio</i>	91
5.35	Principales étapes de traitement avec la méthode <i>diffusion</i>	92
5.36	Principales étapes de traitement avec la méthode <i>intégrée</i>	93
5.37	Transmission numérique de l'information.	95
6.1	Composantes de luminance originales du canal gauche de la deuxième image des séquences testées.	98
6.2	Composante de luminance originale de la deuxième image de la séquen- ce <i>aqua</i>	102
6.3	Composantes de luminance passe-bas de la deuxième image de la sé- quence <i>aqua</i> avec la méthode <i>studio</i>	103

6.4	Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	104
6.5	Composantes de luminance passe-haut agrandies de la région de fixation de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	105
6.6	Composante de luminance reconstruite de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	106
6.7	Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	106
6.8	Composantes de chrominance <i>I</i> de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	107
6.9	Composantes de chrominance <i>Q</i> de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>studio</i>	107
6.10	Signal vidéo composite obtenu avec la méthode <i>studio</i>	108
6.11	Composantes de luminance passe-bas de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i>	110
6.12	Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i>	111
6.13	Composantes de luminance passe-haut agrandies de la région de fixation de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i>	112
6.14	Composante de luminance reconstruite de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i>	113

6.15	Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i> .	113
6.16	Composantes de chrominance <i>I</i> de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i> .	114
6.17	Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>intégrée</i> .	114
6.18	Composantes de luminance passe-haut agrandies de la sous-région de fixation centrale de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>intégrée</i> .	115
6.19	Composantes de luminance passe-haut agrandies de la sous-région de fixation adjacente de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>intégrée</i> .	116
6.20	Composante de luminance reconstruite de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>diffusion</i> .	117
6.21	Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence <i>aqua</i> avec la méthode <i>intégrée</i> .	117
B.1	Géométrie stéréoscopique pour le calcul de la disparité rétinienne	180
C.1	Illustration de la tâche principale	194
C.2	Illustration du mouvement des distracteurs	194

C.3	Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le premier essai des sujets ayant utilisé la stratégie <i>centrale</i> lors de leurs deux essais.	206
C.4	Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le second essai des sujets ayant utilisé la stratégie <i>centrale</i> lors de leurs deux essais.	207
C.5	Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la stratégie <i>centrale</i> lors de leurs deux essais.	207
C.6	Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le premier essai des sujets ayant utilisé la stratégie <i>centrale + droite</i> lors de leurs deux essais.	209
C.7	Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le second essai des sujets ayant utilisé la stratégie <i>centrale + droite</i> lors de leurs deux essais.	209
C.8	Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la stratégie <i>centrale + droite</i> lors de leurs deux essais.	210
D.1	Paramètres pour la spécification d'un filtre.	214
D.2	Filtre passe-bas de 21 coefficients avec une fréquence de coupure normalisée de 0,5.	216

D.3 Réponse fréquentielle du filtre de la figure D.2.	217
D.4 Réponse fréquentielle du filtre de la figure D.2 exprimée en décibels.	217
D.5 Relation de phase de la sous-porteuse f_{sp}	224
D.6 Relation de phase de la sous-porteuse f'_{sp}	225
D.7 Division de l'espace spectral avec la méthode <i>diffusion</i> asymétrique.	228
D.8 Décomposition de l'information de luminance en composantes passe- bas et passe-haut avec la méthode <i>diffusion</i> asymétrique.	229
D.9 Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode <i>diffusion</i>	230
D.10 Reconstruction des composantes à partir du signal vidéo composite avec la méthode <i>diffusion</i> avec trous de Fukinuki.	231
D.11 Séparation des composantes de luminance avec la méthode <i>diffusion</i>	232
D.12 Reconstruction des composantes de luminance passe-haut avec la mé- thode <i>diffusion</i>	233
D.13 Reconstruction des composantes de chrominance du canal gauche avec la méthode <i>diffusion</i>	234
D.14 Reconstruction des composantes de chrominance du canal droit avec la méthode <i>diffusion</i>	234
D.15 Reconstruction des composantes à partir du signal vidéo composite avec la méthode <i>intégrée</i> lorsque la largeur de bande du signal est de 7,16 MHz.	236

D.16 Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode <i>intégrée</i> lorsque la largeur de bande du signal est de 4,2 MHz.	237
D.17 Séparation des composantes de luminance avec la méthode <i>intégrée</i> (première partie).	238
D.18 Reconstruction des composantes de luminance passe-haut avec la méthode <i>intégrée</i> (deuxième partie).	239
D.19 Traitement d'un débordement de la région de fixation vers la gauche de l'image.	242
E.1 Principe de fonctionnement du système de visualisation stéréoscopique.	245

Liste des annexes

A Principes de codage stéréoscopique	167
A.1 Compensation de disparité	167
A.2 Méthodes d'appariement	170
A.2.1 Appariement de caractéristiques	171
A.2.2 Appariement de régions	172
A.3 Comparaison subjective des approches de codage symétriques et asy- métriques	173
A.4 Estimation coopérative de la disparité et du mouvement	175
 B Perception stéréoscopique et attention	 179
B.1 Géométrie de la vision stéréoscopique	179
B.2 Mécanismes de la vision stéréoscopique	183
B.3 Système attentionnel	185
B.4 Système oculo-moteur	187
B.5 Attracteurs d'attention	189

C	Données supplémentaires de l'étude sur les stratégies visuelles . .	192
C.1	Conditions expérimentales	192
C.1.1	Installation	192
C.1.2	Stimuli	193
C.1.3	Méthode	195
C.2	Résultats	196
C.2.1	Direction du mouvement des distracteurs	196
C.2.2	Type de mouvement des distracteurs	198
C.2.3	Distributions des fixations	198
C.2.4	Effet d'apprentissage	204
D	Détails supplémentaires sur les méthodes de codage	211
D.1	Fréquence de modulation	212
D.2	Filtres	213
D.2.1	Filtres spatiaux	213
D.2.2	Filtres temporels	223
D.3	Méthode diffusion asymétrique	226
D.4	Reconstruction avec la méthode diffusion	228
D.5	Reconstruction avec la méthode intégrée	230
D.6	Conditions frontières pour la sélection de la région de fixation	235
E	Données supplémentaires de l'évaluation subjective de la qualité des	

images codées	243
E.1 Conditions expérimentales	243
E.1.1 Système de visualisation	243
E.1.2 Installation	246
E.1.3 Stimuli	247
E.1.4 Méthode	248
E.2 Résultats	250

Introduction

Les systèmes de visualisation stéréoscopique, dont l'utilisation est de plus en plus répandue en multimédia, en téléprésence et en télérobotique, permettent de rehausser la perception tridimensionnelle en présentant à chaque œil un point de vue légèrement différent de la même scène. Les études sur le rôle et l'importance relative des indices visuels de profondeur créés par la stéréoscopie se sont multipliées au cours des dernières années. Une première étude sur le suivi d'un objet se déplaçant en profondeur (Liu, Tharp et Stark 1992) démontre qu'une bonne performance est obtenue en présence d'indices de profondeur stéréoscopique sans indices d'occlusion (représentation de type filiforme), ou en présence d'indices d'occlusion sans indice de profondeur stéréoscopique (représentation polygonale). Toutefois, la meilleure performance est obtenue lorsque les deux indices sont présents conjointement et la pire en l'absence des deux indices. Une autre étude portant sur la manipulation d'objets en profondeur démontre que la présence d'indices de profondeur stéréoscopique, de parallaxe de mouvement ou d'information tactile de retour de force améliorent la performance, alors que la possibilité de faire tourner un objet sur lui-même n'a pas d'effet

significatif sur cette performance (Hirose, Hirota et Kijima 1992). Une étude sur le jugement des relations spatiales dans des images présentées en perspective indique que la visualisation stéréoscopique des images améliore le jugement des relations spatiales tridimensionnelles, et que cette amélioration est d'autant plus grande que les indices de profondeur monoculaires permettant d'inférer les caractéristiques spatiales de la scène sont moins efficaces ou ambigus (Yeh et Silverstein 1992). L'étude présentée par Hendrix et Barfield (1995) démontre que le niveau de présence ressenti dans un environnement virtuel augmente avec la présence d'indices de profondeur stéréoscopique, d'information sonore spatialement localisée et d'indices de parallaxe de mouvement. Toutes ces études, même si elles n'établissent pas catégoriquement la supériorité des indices de profondeur stéréoscopique sur d'autres indices visuels, indiquent clairement que l'utilisation conjointe de la stéréoscopie avec d'autres indices améliore de façon significative la performance des sujets lors de l'accomplissement de diverses tâches. De plus, la présence d'information tactile de retour de force, d'information sonore spatialement localisée et d'indices de parallaxe de mouvement ont aussi des effets positifs sur l'accomplissement de certaines tâches.

Comme l'utilisation de systèmes stéréoscopiques implique une augmentation significative de la quantité d'information visuelle à transmettre à distance, il est essentiel de pouvoir disposer de représentations compactes de cette information. Ces représentations devraient d'une part exploiter toutes les sources de redondance présentes dans les séquences stéréoscopiques et, d'autre part, n'expliciter que l'information

strictement nécessaire au destinataire, l'œil humain dans notre cas. En ce qui concerne tout d'abord les signaux stéréoscopiques, ils possèdent de la redondance intra-image provenant de la cohésion et de la douceur des objets constituant la scène observée, de la redondance temporelle causée par la continuité de mouvement des objets de cette scène et de la redondance stéréoscopique due à l'observation de la même scène à partir de deux points de vue légèrement différents. En ce qui concerne le destinataire, le système visuel humain possède des particularités et des limitations de la perception binoculaire dont la prise en compte permet de comprimer les représentations au delà des limites imposées par la théorie des signaux. Le chapitre 1 présente une revue bibliographique d'approches existantes de compression de séquences stéréoscopiques basés sur ces principes. La majorité des méthodes présentées dépendent d'un appariement local entre les images de chaque paire stéréoscopique. Elles sont donc confrontées aux problèmes inhérents à l'appariement stéréoscopique : sa nature combinatoire, sa susceptibilité aux variations d'éclairage et aux déformations de perspective entre les deux points de vue, sa sensibilité à la présence d'occlusion et son temps de calcul prohibitif.

Le succès d'une méthode de codage ne dépend pas seulement de son efficacité, mais aussi de sa compatibilité avec les systèmes existants. Afin être adoptés dans des domaines d'application utilisant déjà des systèmes d'imagerie monoculaire, les systèmes stéréoscopiques devraient pouvoir être véhiculés par les canaux de communication déjà en place. La plupart des méthodes de codage stéréoscopique présentées

dans la littérature sont destinées à une transmission numérique, et en conséquence, celles voulant préserver une compatibilité monoculaire s'inspirent principalement du standard vidéo MPEG. Il existe pourtant plusieurs domaines d'application où l'information visuelle est transmise de manière analogique, comme la télérobotique, pour lesquels l'utilisation d'information stéréoscopique serait avantageuse. Comme une vaste gamme de systèmes de transmission d'images monoculaires sont basés sur le standard vidéo NTSC, il est raisonnable de tenter de définir des représentations de l'information stéréoscopique qui soient compatibles avec le format NTSC. Ces systèmes pourraient certainement accommoder un format d'image stéréoscopique compatible NTSC. Une solution technologique consistant à transmettre l'information du point de vue gauche dans les champs pairs du format NTSC et celle du point de vue droit dans les champs impairs est d'ailleurs déjà disponible commercialement. Toutefois cette approche peut certainement être améliorée car elle implique des dégradations significatives de la qualité des images dans les directions verticales et temporelles.

La majorité des méthodes de codage stéréoscopiques existantes transmettent l'information visuelle avec le même niveau de détails pour toute l'image. Pourtant, la structure du système visuel humain utilise des résolutions spatiale et de profondeur plus élevées au centre du champ visuel et diminuant rapidement avec une augmentation de l'excentricité rétinienne. Cet aspect du système humain présente un potentiel important de compression car, en principe, il n'est pas nécessaire de transmettre l'information détaillée dans tout le champ visuel, mais seulement à l'intérieur et autour

d'une région de fixation de taille limitée.

Nous avons choisi dans la cadre de ce travail de développer une approche de compression stéréoscopique dont les principales caractéristiques sont : 1) l'absence d'appariement entre les canaux stéréoscopiques, 2) l'exploitation d'une région de fixation, et 3) une compatibilité spectrale avec les systèmes de communication monoculaire. Notre volonté d'éviter une méthode basée sur un calcul explicite de la disparité provient en partie des problèmes inhérents aux méthodes d'appariement stéréoscopique, mais surtout de l'observation que la plupart des méthodes de codage stéréoscopique sont destinées à une observation humaine. Puisque le cerveau humain est le meilleur système fonctionnel connu pour appairer des images et inférer l'information de profondeur à partir de celles-ci, il est donc légitime de lui laisser la tâche d'appariement plutôt que d'investir d'importantes ressources de calcul pour arriver à un résultat moins fiable que celui fourni par la stéréopsie humaine.

L'exploitation de la présence d'une région de fixation avec des images stéréoscopiques ne requiert pas nécessairement d'appariement explicite entre les images. De plus, l'exploitation de cette propriété offre la possibilité de faire une contribution originale dans domaine de la compression d'images parce qu'elle y a été relativement peu utilisée, principalement en raison de la difficulté des deux problèmes fondamentaux qu'elle soulève, soit l'estimation d'une région de fixation et la modulation de la fidélité du codage en fonction de cette région. Ces deux problèmes sont abordés dans ce travail. Le chapitre 2 justifie l'utilisation d'une région de fixation dans le domaine

du codage et présente un aperçu des approches d'estimation de région de fixation présentées dans la littérature. Parmi les approches possibles d'estimation, celle consistant à étudier *a priori* les stratégies visuelles utilisées lors de l'accomplissement d'une tâche a été retenue. A cause du faible nombre d'études accomplies jusqu'ici sur ce sujet pour des conditions de visualisation stéréoscopique, nous avons réalisé une étude psychophysique portant sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur. Cette étude sur les mécanismes de fixation du système visuel humain, qui est décrite au chapitre 4, représente une autre contribution originale de ce travail.

L'approche originale de compression que nous proposons compacte l'information stéréoscopique dans l'espace spectral d'un canal vidéo monoculaire afin de préserver une compatibilité spectrale stéréoscopique/monoculaire. Elle consiste à décomposer l'information stéréoscopique en composantes compactes qui sont par la suite combinées à l'aide de manipulations spectrales afin de former un signal vidéo composite dont l'occupation spectrale n'excède pas celle d'un canal vidéo monoculaire. Il a déjà été reconnu (Isnardi 1988) que le spectre spatio-temporel du signal NTSC pourrait être utilisé plus efficacement et que la qualité du signal NTSC pourrait être significativement améliorée en accomplissant les opérations de filtrage appropriées sur les composantes avant de les positionner dans l'espace spectral pour former un signal vidéo composite. Cette possibilité est aussi implicitement reconnue par Dubois et Schreiber (1988) qui montrent que les filtres numériques représentent des outils tout

à fait adéquats pour mettre en œuvre cette idée. Notre approche suit ces principes en décomposant tout d'abord, à l'aide de filtres numériques à réponse impulsionnelle finie (RIF), les champs gauche et droit de chaque paire stéréoscopique en des composantes de luminance passe-bas et passe-haut. Les composantes de hautes-fréquences sont par la suite limitées à une région de fixation, permettant ainsi une réduction de l'espace spectral qu'elles occupent. La sélection de la région de fixation pour chaque canal est faite en fonction des résultats de notre étude psychophysique, de façon à maximiser la probabilité que le point de fixation de l'observateur se trouve effectivement dans la région de fixation estimée. Un signal vidéo composite est alors formé en positionnant les différentes composantes dans l'espace spectral disponible à l'aide d'opérations de modulation et de filtrage. Les trous de Fukinuki (Fukinuki et Hirano 1984) sont exploités pour positionner les composantes de chrominance des deux canaux stéréoscopiques dans différents quadrants du plan fréquentiel temporel-vertical. L'approche proposée est aussi compatible avec le standard NTSC dans la mesure où elle utilise la même sous-porteuse de chrominance et la même région spectrale pour les composantes de chrominance. Il est donc possible d'utiliser des décodeurs NTSC standard pour séparer les composantes de luminance de celles de chrominance. Un traitement additionnel doit par la suite être accompli afin de récupérer l'information correspondant aux canaux gauche et droit. L'approche proposée est plus efficace, en terme de quantité d'information véhiculée, que celle utilisée par les systèmes commerciaux stéréoscopiques basés sur le standard NTSC, puisque l'information stéréoscopique des

deux canaux est transmise à chaque champ.

L'exploitation de la non-uniformité des résolutions spatiales et stéréoscopiques du système visuel humain consiste en le fait de ne pas transmettre l'information de haute-fréquences dans tout le champ visuel, mais seulement à l'intérieur d'un sous-espace contenant la région de fixation. L'information transmise est ainsi représentée plus finement dans la région de fixation qu'en périphérie. La résolution stéréoscopique est également implicitement réduite à l'extérieur de la région de fixation puisque la finesse de la correspondance entre les canaux gauche et droite y est réduite. Cet aspect du travail représente une contribution originale car l'exploitation d'une région de fixation n'a jamais été appliqué à la compression spectrale de séquences stéréoscopiques. Un autre aspect innovateur de ce travail provient de l'intégration aux méthodes de codage de certains résultats de l'étude du chapitre 4 sur les processus de fixation du système visuel humain, une démarche qui illustre à la fois la pertinence de notre investigation psychophysique et qui augmente la crédibilité des méthodes de codage qui sont proposées.

Trois méthodes de codage, toutes basées sur les principes qui viennent d'être décrits, sont proposées afin d'exploiter pleinement la bande passante de différents systèmes de communication : (1) la méthode *studio* pour laquelle la limite spectrale horizontale correspond à la moitié de la fréquence d'échantillonnage des images originales, soit 7.16 MHz, (2) la méthode *diffusion* pour laquelle la limite spectrale horizontale est fixée à 4,2 MHz, et (3) la méthode *intégrée* qui peut satisfaire si-

multanément les contraintes de bande passante des deux premières méthodes. Le format *studio* est approprié, entre autres, pour une utilisation avec des enregistreuses numériques D2 pour la vidéo composite, alors que le format *diffusion* s'applique surtout à la transmission d'images.

La thèse est organisée de la façon suivante : le chapitre 1 présente une revue des approches de codage stéréoscopique, le chapitre 2 justifie l'exploitation d'une région de fixation en codage, le chapitre 3 donne un aperçu plus global des méthodes de codage et présente le système général de transmission d'images dans lequel elles s'insèrent et le chapitre 4 décrit l'étude sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur. Les méthodes de codage proposées sont décrites en détails au chapitre 5. Le chapitre 6 présente ensuite des résultats expérimentaux illustrant la validité des méthodes proposées. Une brève conclusion résume ce travail, discute des améliorations possibles à y apporter et présente certaines questions qui demeurent ouvertes et pourraient éventuellement faire l'objet de travaux de recherche futurs.

Chapitre 1

Un survol bibliographique des approches de codage stéréoscopique

Une méthode de codage stéréoscopique devrait idéalement exploiter toutes les sources de redondance des signaux stéréoscopiques. Ces signaux possèdent une redondance intra-image provenant de la douceur et de la cohésion des objets d'une scène. Les techniques de codage pour images fixes, telles que le codage prédictif ou le codage par transformée (Jain 1989), exploitent la redondance intra-image en profitant du fait que la majorité des scènes sont constituées de surfaces localement continues et douces qui se projettent en des régions d'image de brillance homogène ou à variations douces séparées par des discontinuités, et qui contiennent donc principalement de

l'information passe-bas. Dans le cas de scènes dynamiques, la continuité des mouvements des objets génère une redondance temporelle (inter-images) dont l'élimination peut donner lieu à une réduction importante de l'information à transmettre. La compensation de mouvement représente la principale stratégie de codage de séquences d'images et consiste à transmettre un champ de vecteurs de déplacement qui, appliqué à l'image précédente, permet la reconstruction de l'image courante (Musmann, Pirsh et Grallert 1985). Un vecteur de déplacement par région de forme et de taille pré-établies est généralement estimé pour chaque bloc dans l'image originale. Chaque région de l'image reconstruite est par la suite obtenue en la remplaçant par la région correspondante de l'image précédente. En plus des vecteurs de déplacement, les erreurs de prédiction (c'est-à-dire les écarts entre les brillances réelles et les brillances prédites) adéquatement quantifiées, peuvent aussi être transmises afin d'améliorer la qualité des images reconstruites.

Les signaux stéréoscopiques possèdent une redondance supplémentaire engendrée par l'observation d'une même scène à partir de deux points de vue légèrement différents. La majeure partie de la scène est donc visible des deux points de vue et se projette en des régions similaires sur les plans image gauche et droit. Cette redondance géométrique peut donc en théorie être exploitée à des fins de compression. Une stratégie possible consiste à mettre en correspondance des régions de forme et de taille pré-établies entre les deux images d'une paire stéréoscopique et de transmettre les vecteurs de disparité permettant la reconstruction d'un point de vue à partir de

l'autre. Cette stratégie de prédiction est appelée *compensation de disparité* (Ozkan et Salari 1993, Seferidis et Papadimitriou 1993, Dinstein, Guy, Rabany, Tzelgov et Henik 1988). L'annexe A présente en détails le principe de compensation de disparité ainsi qu'un résumé des principales approches d'appariement stéréoscopique.

Les méthodes de compression stéréoscopique peuvent aussi exploiter les caractéristiques perceptives du système visuel humain reliées à la stéréoscopie. Par exemple, les limites de la résolution de profondeur du système visuel humain peuvent être prises en compte pour limiter l'amplitude des vecteurs de disparité à des valeurs supérieures au seuil de discrimination de profondeur (Steurer 1992, Yamaguchi, Tatehira, Akiyama et Kobayashi 1989). Il est aussi possible que certains systèmes de présentation d'images stéréoscopiques ne puissent reproduire toute la gamme des valeurs de disparité perceptibles par le système visuel humain. Dans ces cas, la quantification des vecteurs de disparité peut se faire en fonction des limites du système de présentation (Yamaguchi et al. 1989, Kost 1990). Le fait que la résolution de profondeur du système visuel humain décroisse de façon exponentielle avec une augmentation de la distance du plan de fixation peut aussi être exploité afin de représenter les vecteurs de disparité avec une résolution non-uniforme selon une fonction de quantification exponentielle (Schertz 1992, Schertz, Fayad et Bodamer 1991). Cette approche suppose que le plan de fixation des observateurs est toujours placé sur l'écran du système de visualisation, une hypothèse qui, selon nos résultats expérimentaux du chapitre 4, n'est que partiellement valide. La décroissance des résolutions spatiale et stéréoscopique

en fonction d'une augmentation de l'excentricité rétinienne peut aussi être prise en compte afin de réduire le niveau de détails des images dans les régions périphériques (Obrador 1995). Puisque cette approche de compression est utilisée dans le cadre de ce projet, ses avantages, ses exigences ainsi que ses limites seront présentées plus en détails au chapitre 3.

Le seuil de perception des fréquences spatiales de disparité peut être utilisé pour limiter la largeur de bande de l'information de disparité à transmettre (Steurer 1992, Kost 1990). Il est aussi possible de se baser sur le seuil de perception de différences d'intensité entre les canaux gauche et droit pour quantifier l'erreur de prédiction transmise avec les vecteurs de disparité (Steurer 1992). Finalement, une quantification adaptative des vecteurs de disparité peut se faire avec comme objectif une réduction de la perception d'artefacts visuels tridimensionnels en tenant compte d'indicateurs ayant été spécifiquement établis en fonction d'un visionnement stéréoscopique (Tseng et Anastassiou 1995). Les deux principales théories de la vision stéréoscopique humaine, la théorie de la suppression et celle de la fusion, peuvent aussi être mises à contribution pour l'élaboration de méthodes de codage, comme le décrivent les sections suivantes.

1.1 Théorie de la suppression et codage asymétrique

Selon la théorie de la suppression (ou neutralisation), le percept tridimensionnel subjectif résultant d'une paire stéréoscopique avec une image dégradée possède un niveau de qualité s'approchant de celui de l'image non dégradée (Arditi 1986). L'image dégradée contribue au calcul de la disparité puis est *supprimée* pour être remplacée par l'autre image. Cette théorie inspire les approches codage dites *asymétriques*, où les deux images d'une paire stéréoscopique sont traitées différemment, généralement en utilisant une image comme une référence pour la prédiction de l'autre. Les approches de codage asymétrique prennent souvent pour acquis que les détériorations résultant de la compression élevée d'une des deux images de la paire sont peu perceptibles parce que l'autre image de la même paire conserve un bon niveau de qualité (Perkins 1992, Zhu, Gao et Goutte 1991, Tzelgov, Henik et Dinstein 1990). Par exemple, il existe des méthodes qui exploitent la théorie de la suppression en sous-échantillonnant une des deux images d'une paire stéréoscopique (Perkins 1992, Dinstein, Kim et Henik 1991, Zhu et al. 1991, Dinstein, Kim, Tselgov et Henik 1989). Même si ces méthodes de codage sont attrayantes car en général elles ne requièrent pas d'appariement entre les canaux gauche et droit, elles peuvent s'avérer inadéquates pour des applications nécessitant la perception de détails fins (Perkins 1992), comme la télémanipulation par exemple. Le codage par compensation de disparité représente un autre exemple

typique de codage asymétrique (voir l'annexe A). Il est aussi possible d'exploiter d'autres relations géométriques que celles concernant la disparité pour la compression d'images stéréoscopiques. Par exemple, l'information concernant les transformations projectives permettant la reconstruction d'un point de vue à partir de l'autre peut être transmise (Oisel, Morin et Labit 1995).

1.2 Théorie de la fusion et codage symétrique

Selon la théorie de la fusion, les deux images d'une paire stéréoscopique sont fusionnées pour donner lieu à une représentation volumétrique de l'environnement (Arditi 1986). Cette théorie motive les approches de codage symétrique impliquant la création d'une représentation intermédiaire intégrant l'information des deux images d'une paire stéréoscopique. Avec ces approches, les erreurs de reconstruction sont généralement distribuées également entre les deux images. La représentation intermédiaire peut prendre diverses formes, comme par exemple une autre image obtenue en combinant les deux images stéréoscopiques originales ou un modèle tridimensionnel intégrant l'information des deux points de vue.

Une image intermédiaire correspondant à l'image qui serait obtenue d'une caméra positionnée au centre d'un système stéréoscopique peut être générée à partir du champ de disparité entre les deux images stéréoscopiques originales. L'intensité de chaque pixel de cette image est calculée en faisant la moyenne des intensités des pixels correspondants dans les images gauche et droite (De Vleeschauwer 1991a).

Une autre méthode pour l'obtention d'une image intermédiaire consiste à multiplexer temporellement les images gauche et droite d'une séquence stéréoscopique de façon adaptative en tenant compte de l'amplitude des erreurs de reconstruction provenant de prédictions par compensation de disparité et de mouvement (Chassaing et Pelé 1992, Chassaing et Pelé 1991, Chassaing, Choquet et Pelé 1991). Afin de maintenir la consistance de l'image intermédiaire, un des deux points de vue doit être décalé en fonction de la disparité.

Les approches de codage symétrique et asymétrique ont été comparées de façon subjective dans (Tzelgov et al. 1990, Dinstein, Guy, Rabany, Tzelgov et Henik 1989, Dinstein et al. 1988). Les principaux résultats de cette comparaison sont présentés à l'annexe A.

1.3 Exploitation simultanée du mouvement et de la stéréoscopie

Dans le cas de séquences stéréoscopiques, plusieurs sources de redondance sont simultanément présentes : la structure intra-image, le mouvement et la stéréoscopie. Des approches de compression adaptatives peuvent être développées afin d'exploiter pleinement chacune de ces sources. Plusieurs méthodes, s'inspirant du standard MPEG développé pour la compression de séquences monoculaires, ont été proposées afin de prendre aussi en compte la redondance stéréoscopique. En général, un module

de sélection du mode de prédiction est utilisé pour l'encodage des blocs d'image en tenant compte des erreurs de reconstruction au récepteur. Le canal gauche est encodé au moyen d'une transformée en cosinus discrète (TCD) et d'une compensation de mouvement, et le canal droit par compensation de mouvement, compensation de disparité ou encodage direct du bloc image avec TCD, en choisissant pour chaque bloc l'estimé donnant lieu à la plus petite erreur de reconstruction (Chassaing 1995, Chassaing et Pelé 1993, Kopernik et Pelé 1993b, Kopernik et Pelé 1992, Pelé, Chassaing et Kopernik 1992, Horst 1992). La compensation de disparité et de mouvement peut aussi être appliquée à une image intermédiaire obtenue en multiplexant le point de vue gauche avec le point de vue droit décalé (Chassaing et Pelé 1992, Chassaing et Pelé 1991, Chassaing et al. 1991). Afin d'améliorer la performance de ces méthodes, les champs de mouvement et de disparité devraient être estimés de façon consistante afin que les différents modes de prédictions puissent être utilisés de façon flexible. L'estimation coopérative du mouvement et de la disparité, décrite à l'annexe A, permet d'obtenir cette consistance.

Afin d'éviter les artefacts inhérents aux approches utilisant des blocs d'image lorsque des taux de compression élevés sont appliqués, il est possible de traiter l'information d'une image en terme d'objets bidimensionnels (Ziegler et Panis 1995, Ziegler, Panis et Cosmas 1993). Une procédure itérative utilisant l'information stéréoscopique disponible est utilisée pour segmenter l'image en régions uniformes correspondant à des objets (Ziegler et al. 1993). L'information de mouvement peut aussi être prise en

compte lors de la segmentation (Ziegler et Panis 1995). Les objets en mouvement sont transmis en utilisant un ensemble de paramètres décrivant leur forme, leur couleur, leur mouvement et leur disparité. Pour les régions pour lesquelles des modèles d'objets ne peuvent être évalués de façon satisfaisante, la différence temporelle entre les images successives est transmise de façon conventionnelle.

Les informations dynamique et stéréoscopique peuvent être intégrées en une représentation par objets compacte contenant les paramètres de structure 3D et de mouvement 3D (Tzovaras, Grammalidis et Strintzis 1995, Grammalidis, Malassiotis, Tzovaras et Strintzis 1993, Dugelay 1992, Dugelay et Pelé 1992, Morikawa et Harashima 1991). La modélisation par objet peut faciliter l'estimation du mouvement et de la disparité en tenant compte, par exemple, de la différence entre l'image courante et la projection de l'estimé du modèle 3D sur cette image (Falkenhagen 1995). L'information transmise peut être représentée en terme d'objets 3D constitués de facettes planaires dont le mouvement est représenté par une composante de rotation et une composante de translation (Dugelay et Pelé 1992). La surface 3D peut aussi être représentée par l'information de profondeur aux passages par zéros, obtenus d'un détecteur de contour (Morikawa et Harashima 1991). Chaque objet peut être représenté par son mouvement 3D, sa structure 3D et ses paramètres de texture (Falkenhagen 1995). La forme des objets est représentée par un maillage triangulaire 3D obtenu en projetant les maillages triangulaires 2D de régions segmentées de chaque canal stéréoscopique sur la carte de profondeur calculée à partir

de l'information stéréoscopique et dynamique.

Même si la modélisation tridimensionnelle d'une scène représente une approche attrayante pour le codage de séquences stéréoscopiques, la complexité de cette modélisation est souvent très importante. Il est aussi possible que cette modélisation soit inadéquate pour certaines régions des images, en raison d'occlusions par exemple. Un avantage de l'extraction explicite de l'information de profondeur est que les objets frontaux peuvent facilement être séparés du fond de la scène (Waldowski 1991). Il est alors possible de traiter différemment ces deux entités en utilisant des approches hybrides combinant des représentations 2D et 3D. La sélection d'un mode *objet* (3D) ou d'un mode *bloc* (2D) peut se faire en fonction d'un objectif de minimisation des erreurs de reconstruction (Grammalidis et al. 1993). Une autre possibilité consiste à représenter les objets ayant d'importantes variations de profondeur avec une information 3D et ceux avec de faibles valeurs de profondeur avec une information 2D (Kopernik et Pelé 1993a, Kopernik et Pelé 1993b).

En résumé, la majorité des approches de codage stéréoscopique présentées dans ce chapitre ont en commun une étape d'appariement qui permet d'obtenir un champ de disparité à partir duquel est accompli soit une prédiction par compensation de disparité ou soit le calcul d'une représentation intermédiaire. Dans le cas de séquences stéréoscopiques, l'information additionnelle apportée par le mouvement permet le développement des schémas de prédiction adaptatifs et facilite la représentation d'une scène en terme d'objets tridimensionnels. Pour les approches utilisant une prédiction

par compensation de disparité, plusieurs propriétés du système de vision stéréoscopiques peuvent être prises en compte pour la quantification des vecteurs de disparité. La théorie de la suppression du système visuel humain peut aussi être exploitée afin de réduire le niveau de détails d'un des deux canaux stéréoscopiques.

Chapitre 2

L'exploitation d'une région de fixation en codage

Toute personne avec une vision normale peut facilement constater que, pour l'observation d'une scène avec un seul œil dans des conditions d'éclairage diurne, il est plus facile de percevoir les détails fins dans le centre du champ visuel qu'en périphérie. La précision avec laquelle les détails spatiaux sont détectés et identifiés est appelée l'*acuité spatiale*. Des études psychophysiques démontrent que l'acuité spatiale dépend de la position des objets dans le champ visuel et diminue de façon monotone avec une augmentation de leur distance du point de fixation (Olzak et Thomas 1986). L'explication de ce phénomène provient de la structure anatomique du système visuel humain qui possède une densité des récepteurs pour la vision de jour (cônes) moins élevée en périphérie qu'au centre (Olzak et Thomas 1986). Lors de l'observa-

tion d'une scène avec les deux yeux, la stéréoscopie peut être utilisée pour obtenir de l'information sur la profondeur de cette scène. La précision avec laquelle l'information de profondeur est résolue est appelée l'*acuité stéréoscopique*. De la même façon que l'acuité spatiale, l'acuité stéréoscopique diminue avec une augmentation de l'excentricité rétinienne. Plusieurs facteurs affectent la valeur du seuil d'acuité stéréoscopique, dont la taille des stimuli, l'excentricité rétinienne et la présence de stimuli adjacents. En particulier, l'acuité stéréoscopique est maximale à la fovéa et diminue avec une augmentation de l'excentricité rétinienne, et de façon plus rapide après 6 deg (Arditi 1986). Les seuils d'acuité stéréoscopique les plus petits se situent au plan de fixation des yeux et ils augmentent avec une augmentation de la disparité, qu'elle soit croisée ou non-croisée (Arditi 1986). A titre indicatif, le tableau B.1 de l'annexe B présente des valeurs représentatives de limites de disparité pour deux tailles de stimuli et deux valeurs d'excentricité rétinienne. L'annexe B présente aussi la géométrie stéréoscopique impliquée dans la définition d'un point de fixation dans l'espace, ainsi qu'un résumé du fonctionnement de mécanismes de base de la vision stéréoscopique associés à la perception de profondeur.

La tendance naturelle du système visuel humain consiste à bouger les yeux afin que les objets d'intérêt d'une scène se projettent sur la fovéa pour un examen en détails. L'attention sélective est souvent considérée comme le mécanisme contrôlant le mouvement des yeux de façon à diriger leur fixation vers un objet d'intérêt. Par contre, il existe des situations où l'attention n'est pas dirigée vers un objet étant projeté sur la

fovéa. Par exemple, dans (Faubert et Von Grünau 1995), des phénomènes d'attention divisée sont rapportés, ce qui implique la présence simultanée de plus d'une région d'intérêt. Comme il est impossible de fixer simultanément plus d'un endroit dans l'espace, il est clair qu'une distinction doit être faite entre la notion de région d'intérêt et celle de région de fixation. Toutefois, le couplage étroit entre le système attentionnel et le système oculo-moteur entraîne souvent une confusion quant au rôle respectif de chaque système. La principale conséquence de cette confusion est que les notions de région d'intérêt et de région de fixation sont souvent utilisées indistinctement. Dans un contexte de codage, seule la notion de région de fixation est pertinente, puisque la résolution avec laquelle les détails d'une image sont perçus est fonction de la position du point de fixation des yeux. L'annexe B présente une revue des principales théories sur le système attentionnel qui tentent d'expliquer les mécanismes de sélection d'une région d'intérêt, ainsi qu'un aperçu des caractéristiques du système oculo-moteur dont le rôle est de contrôler les changements de position de la région de fixation.

L'utilisation d'une région de fixation est prometteuse pour le codage de séquences stéréoscopiques puisque, une fois la position du point de fixation connue, une compression de l'information à transmettre est possible en modulant le niveau de détails des images en fonction des acuités spatiales et stéréoscopiques du système visuel humain. Ainsi, l'information de hautes-fréquences peut être limitée à une région de fixation et ne doit donc pas être transmise en périphérie. Le succès de cette approche repose toutefois sur une estimation fiable et adéquate de la région de fixation. Cette

estimation est un problème difficile car elle implique des interactions complexes entre des processus cognitifs de haut-niveau et des routines visuelles de plus bas niveau contrôlant les mouvements des yeux. L'approche d'estimation qui vient naturellement à l'esprit consiste à prédire la position de la région de fixation en fonction de caractéristiques des images qui capturent l'attention et attirent les mouvements des yeux. Plusieurs caractéristiques des images, énumérées à l'annexe B, ont été établies comme des attracteurs de fixation dans certaines situations. Toutefois, les résultats de Yarbus (1967), entre autres, mettent en doute l'efficacité d'une approche d'estimation basée sur le contenu des images en démontrant que, pour une même image, la séquence des points de fixation peut changer de façon significative en fonction de la tâche des sujets. En conséquence, l'estimation d'une région de fixation doit généralement se faire en fonction d'une tâche ou d'un contexte spécifique. La majorité des approches d'estimation de région de fixation rapportées dans la littérature s'inscrivent dans ce cadre en proposant des critères de sélection de région de fixation en fonction de tâches spécifiques ou d'objectifs précis. Ces critères peuvent être inspirés des propriétés du système visuel humain (voir la section B.5) ou simplement conçus afin de faciliter l'interprétation d'information visuelle, pour des applications de vision informatique par exemple (Abbott 1992).

Dans le cas de tâches d'inspection, la région de fixation peut être fixe et connue *a priori*. Pour des applications de surveillance, les régions contenant des objets en mouvement peuvent être choisies comme régions de fixation (Nguyen, Labit et

Odobez 1994, Burt 1988). Lors de la détection d'objets, une carte de saillance établie en fonction des caractéristiques de bas niveau d'une image peut être utilisée afin de déterminer la région de fixation (Clark et Ferrier 1988). L'inférence de la structure tridimensionnelle d'une scène, l'estimation de surface et la modélisation volumétrique peuvent être facilitées en choisissant le prochain point de fixation en fonction de la distance stéréoscopique estimée (Krotkov 1987), de la proximité des points de la caméra (Ahuja et Abott 1993), ou d'une réduction espérée de l'incertitude (Shmuel et Werman 1990, Whaite et Ferrie 1990). Pour la reconnaissance d'objets, la surface qui permet le mieux de désambigüer l'hypothèse courante sur l'identité du candidat peut être choisie comme prochain point de fixation (Grimson 1986). L'information de plus basse résolution se trouvant en périphérie peut aussi être choisie comme futur point de fixation afin d'aider à établir l'identité d'un objet (Browse et Rodrigues 1987). Afin de faciliter la reconnaissance d'objets, il peut aussi être avantageux de fixer les caméras sur un objet de façon à le décrire en fonction d'un cadre de référence établi en fonction de cet objet (Ballard 1990). De cette façon un modèle de projection orthographique peut être utilisé pour les caractéristiques de l'image se trouvant à proximité de l'axe optique, simplifiant ainsi grandement l'analyse mathématique nécessaire à la reconnaissance. Pour des applications de navigation autonome et de reconnaissance d'objet, le choix du prochain point de fixation peut se faire de façon à découvrir des régions d'image jusque là inconnues ou inexplorées (Califano, Kjeldsen et Bolle 1990, Xie 1990).

Lorsque les images sont directement présentées à un observateur humain, il est possible d'accomplir un suivi continu du mouvement des yeux. Pour des applications de téléopération, des systèmes de mesure du mouvement des yeux sont utilisés afin d'obtenir directement les points de fixation de l'opérateur (Obrador 1995). Des appareils de mesure du mouvement des yeux sont aussi utilisés avec des systèmes de télévision tridimensionnelle afin d'augmenter la portion de l'image où l'information stéréoscopique détaillée est présentée (Tetsutani, Kishino et Nagashima 1992). Les méthodes basées sur la mesure des mouvements oculaires doivent fonctionner en temps réel et requièrent souvent un équipement encombrant qui peut à la longue fatiguer l'utilisateur. Lorsque de telles méthodes ne peuvent être utilisées, il est aussi possible de prédire la région de fixation directement à partir du contenu des images associées à une tâche spécifique. Dans ce cas, l'estimation de la région de fixation doit être basée sur une étude préalable du mouvement des yeux des observateurs. Par exemple, une forte corrélation entre les points de fixation des sujets est observée lors du visionnement de séquences vidéo représentatives d'émission de télévision typiques (Stelmach, Tam et Hearty 1991). Une autre étude portant sur les mouvements des yeux lors d'une tâche de recherche visuelle rapporte que différents types de séquences de fixation sont observés selon la nature et la complexités des images dans lesquelles la recherche est accomplie (Stark, Ezumi, Nguyen, Paul, Tharp et Yamashita 1992). Lors de cette étude, la recherche se déroulait dans des images 2D denses avec des cibles dispersées aléatoirement et dans des images 3D naturelles avec des cibles placées à

des endroits réalistes. Une stratégie de recherche régulière consistant en une séquence de mouvements des yeux répétitifs et systématiques permettant de balayer une région entière de l'image a été observée lorsque la recherche se faisait sur de petites régions des images 2D à l'aide de la vision centrale. Lorsque l'espace de recherche était agrandi de façon à ce que la vision périphérique puisse jouer un rôle dans la recherche, une stratégie de recherche irrégulière était alors utilisée. Dans le cas de scènes naturelles contenant des objets connus, un apprentissage des endroits possibles où pouvaient se trouver les cibles avait présumément déjà eu lieu puisque ce sont surtout ces endroits qui ont été examinés. Cette étude ainsi que celle de (Stelmach et al. 1991) suggèrent l'existence de comportement de visionnement spécifiques à certaines tâches visuelles.

En résumé, les mécanismes de fixations du système visuel humain permettent de percevoir de façon détaillée les objets d'intérêt d'une scène sans toutefois nécessiter une représentation simultanée de toutes les entités de cette scène à un niveau de détails élevé. Des méthodes de codage s'inspirant des processus de fixation du système visuel humain peuvent donner lieu à une compression importante de l'information à transmettre puisque les détails fins n'ont pas à être représentés à l'extérieur de la région de fixation. Le succès de ces méthodes repose sur une estimation adéquate et fiable de la région de fixation. Cette estimation doit être accomplie en fonction d'un contexte ou d'une tâche spécifique et peut avoir comme but une facilitation de l'interprétation d'information visuelle pour des applications de vision informatique

et de robotique, ou une reproduction des séquences de fixations d'un observateur pour des applications impliquant une supervision humaine. Pour ces applications, il est aussi possible d'obtenir directement les fixations d'un observateur à partir d'un système de suivi des mouvements oculaires.

Chapitre 3

Un aperçu général de l'approche de codage proposée

3.1 Contexte et justification

Nous avons choisi de favoriser une approche de codage ne nécessitant pas d'appariement entre les deux canaux stéréoscopiques. La majorité des approches de codage stéréoscopique présentées au chapitre 1 sont basées sur un calcul explicite de la disparité, que ce soit pour tenir compte des propriétés perceptives du système de vision stéréoscopique ou pour coder des images stéréoscopiques par compensation de disparité. Ces approches sont nécessairement confrontées à tous les problèmes inhérents aux méthodes d'appariement stéréoscopique, tels que la présence d'occlusions, les différences de conditions d'éclairage entre les canaux gauche et droit, les

déformations de perspective entre les deux points de vue et des temps de calcul souvent prohibitifs. Notre volonté d'éviter une méthode basée sur un calcul explicite de la disparité provient entre autre de l'observation que la plupart des méthodes de codage stéréoscopique sont destinées à une observation humaine et que le cerveau humain est le meilleur système fonctionnel connu pour apparier des images et inférer l'information de profondeur à partir de celles-ci. Il est donc légitime de laisser la tâche d'appariement au système visuel humain plutôt que d'y investir d'importantes ressources de calcul pour arriver à un résultat moins fiable que celui fourni par la perception humaine. Ce choix d'approche élimine *a priori* la majorité des méthodes de codage stéréoscopiques présentées au chapitre 1. Seules les approches s'inspirant de la théorie de la suppression pour réduire le contenu fréquentiel d'un des deux canaux stéréoscopiques ne requièrent pas d'appariement (Dinstein et al. 1991, Dinstein, Kim, Tselgov et Henik 1989, Perkins 1992, Zhu et al. 1991). Toutefois, ces approches peuvent ne pas être appropriées pour des applications nécessitant une perception de détails fins, comme la télérobotique ou la télémanipulation.

La structure du système visuel humain avec des résolutions spatiales et de profondeur plus élevées au centre du champ visuel et qui diminuent rapidement avec une augmentation de l'excentricité rétinienne a été identifiée comme une propriété pouvant donner lieu à une compression importante parce que l'information détaillée ne doit plus être transmise pour toute l'image, mais seulement à l'intérieur et autour d'une région de fixation. L'exploitation de cette propriété avec des images stéréoscopiques

ne requiert pas nécessairement d'appariement explicite entre les images. De plus, l'exploitation de cette propriété offre la possibilité de faire une contribution originale dans le domaine de la compression d'images parce qu'elle y a été relativement peu utilisée, principalement en raison de la difficulté des deux problèmes fondamentaux qu'elle soulève, soit l'estimation d'une région de fixation et la modulation de la fidélité du codage en fonction de cette région.

A la suite de la revue des principales méthodes d'estimation de région de fixation du chapitre 2, il est clair que, compte tenu du rôle prépondérant de l'opérateur humain dans la boucle de contrôle d'un système de télérobotique, toute approche d'estimation dont l'objectif ne consiste pas en une reproduction la plus fidèle possible des séquences de fixation de l'opérateur doit être rejetée. Deux types d'approches d'estimation s'avèrent donc appropriées au contexte particulier de visualisation stéréoscopique pour la télérobotique : l'utilisation d'un système de suivi des mouvements oculaires ou l'exploitation des résultats d'une étude sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche d'intérêt. La première approche d'estimation représente une solution de nature technologique au problème d'estimation, tandis que la seconde suscite des questions fondamentales sur le fonctionnement du système visuel humain. De plus, les quelques études rapportées au chapitre 2 qui s'inscrivent dans le cadre de cette seconde approche suggèrent l'existence de comportements de visionnement spécifiques à certaines tâches qui pourront être exploités pour le développement d'une stratégie d'estimation de région de fixation. A la suite

de ces observations, nous avons choisi d'utiliser pour ce travail une méthode d'estimation de région de fixation basée sur les résultats d'une étude sur les stratégies visuelles d'un opérateur. En se basant sur les arguments apportés au chapitre 2, il est clair que cette étude doit être faite en fonction de l'accomplissement d'une tâche spécifique. Les principaux critères que nous avons identifiés pour la sélection de cette tâche sont :

- 1) qu'elle soit typique de la télérobotique puisque c'est un contexte important d'application de ce travail ;
- 2) qu'elle implique la nécessité d'information stéréoscopique pour la perception de profondeur puisqu'une des principales justifications de ce travail provient de l'apport positif de la stéréoscopie à l'accomplissement de certaines tâches (voir l'introduction) ;
- et 3) qu'elle soit dynamique afin d'être la plus réaliste possible.

L'absence dans la littérature d'études satisfaisant ces critères nous a amené à réaliser notre propre étude psychophysique sur les stratégies visuelles d'un opérateur lors de l'accomplissement d'une tâche de discrimination de profondeur. Cette étude est décrite en détails au chapitre 4.

Le développement d'une méthode de codage doit se faire, entre autres, en fonction du canal de communication par lequel l'information est transmise. Dans le cadre de ce travail, il a été choisi de développer une méthode de compression spectrale pour la transmission d'information stéréoscopique par un canal analogique. Les principales motivations de ce choix ont été présentées à l'introduction. Il est légitime, en fonction de ce choix, de tenter de développer des méthodes de compression qui conservent une certaine compatibilité avec le standard NTSC, puisqu'après tout c'est le format vidéo

le plus répandu en Amérique du Nord. Nous avons donc choisi de nous inspirer de ce standard vidéo pour développer nos méthodes de compression, mais sans toutefois faire de la compatibilité avec ce standard une condition *sine qua non*, et ce afin que l'élaboration de ces méthodes ne se fasse pas qu'en fonction de choix purement technologiques. De façon plus précise, le problème qui est traité dans ce travail peut être formulé de la façon suivante :

Soit $\{Y_g, Y_d\}$ et $\{I_g, Q_g, I_d, Q_d\}$, les composantes de luminance et de chrominance des canaux gauche et droit d'une séquence stéréoscopique entrelacée, échantillonnées respectivement à $4f_{sp}$ et $2f_{sp}$. Etant donné ces composantes, il s'agit de former un signal vidéo composite contenant les deux canaux de la séquence représentés avec une résolution spatiale inhomogène en fonction d'une région de fixation, et occupant un espace spectral n'excédant pas celui d'un signal NTSC. De plus, afin de conserver une compatibilité avec le standard NTSC, on désire utiliser la sous-porteuse de chrominance f_{sp} du standard NTSC pour l'information de chrominance d'au moins un des canaux stéréoscopiques.

Deux types de signaux NTSC sont considérés dans la cadre de ce travail, soit le signal *studio* pour lequel la largeur de bande est égale à $2f_{sp}$ et le signal *diffusion* pour lequel la largeur de bande est limitée à 4,2 MHz. La fréquence de $2f_{sp} = 7,16$ MHz pour le format *studio* semble plus grande que la largeur usuelle de 6 MHz du signal vidéo NTSC. Cependant, elle est justifiable en raison de la disponibilité d'enregistreuses numériques D2 pour la vidéo composite. De son côté le format *diffusion* s'applique

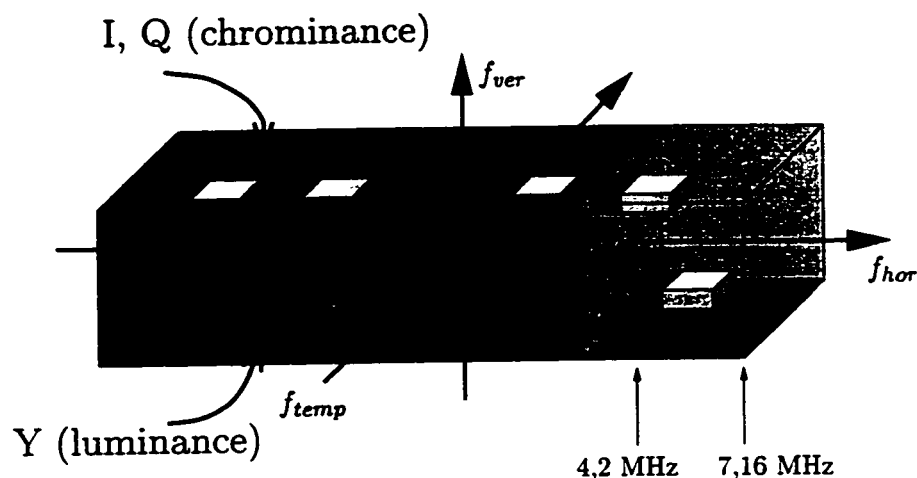


Figure 3.1 : Occupation spectrale d'un champ du signal NTSC dans le volume fréquentiel horizontal-vertical-temporel.

surtout à la transmission d'images. Pour les deux types de signaux, le signal contient 480 lignes actives par image, ce qui limite l'espace spectral vertical à $f_v = 120$ c/hi pour chaque champ. Les composantes de chrominance occupent un espace spectral centré à (f_{sp}, f_v) dans le plan fréquentiel horizontal-vertical et à $(15 \text{ Hz}, f_v)$ dans le plan fréquentiel temporel-vertical. L'espace spectral occupé par les deux types de signaux est illustré aux figures 3.1 et 3.2.

A la figure 3.2.c, la position des trous de Fukinuki (Fukinuki et Hirano 1984) est illustrée. Ces trous ne sont pas exploités par les signaux NTSC standard. Il est néanmoins possible d'y insérer de l'information additionnelle en utilisant une sous-porteuse auxiliaire f'_{sp} dont la fréquence est la même que celle de f_{sp} , mais dont la phase est inversée par rapport à f_{sp} à tous les deux champs.

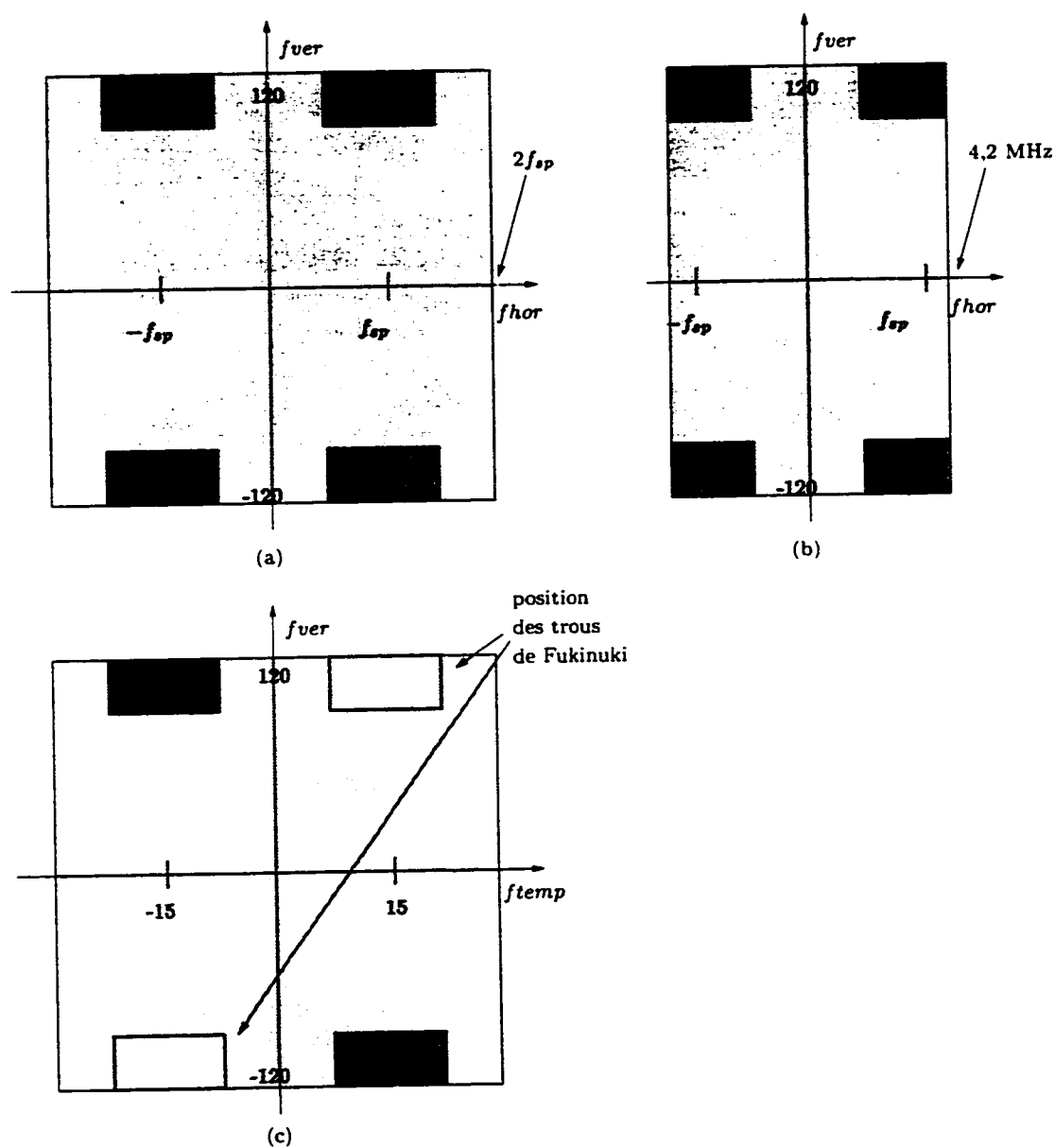


Figure 3.2 : Occupation spectral d'un champ du signal NTSC dans les plans fréquentiels horizontal-vertical et temporel-vertical : plan horizontal-vertical pour (a) un signal *studio* et (b) un signal *diffusion* ; (c) plan temporel-vertical pour les deux signaux.

3.2 Système de transmission d'images

La figure 3.3 présente de le système de transmission d'images à l'intérieur duquel s'inscrit le présent travail. Les images d'une scène sont captées par une paire de caméras stéréoscopiques en couleur. Ces caméras produisent chacune les trois signaux analogiques du standard vidéo NTSC, soit les composantes Y , I et Q en format 4:2:2. Ces signaux sont transformés d'un format analogique à un format numérique à l'aide de convertisseurs analogique/numérique (Oppenheim et Schafer 1989). La compression des signaux image se fait numériquement, puis le signal vidéo composite provenant de la sortie de l'encodeur est mis sous forme analogique à l'aide d'un convertisseur numérique/analogique. La compression des composantes se fait en fonction d'une région de fixation dont la position est fournie par le module d'estimation de région de fixation. Le signal vidéo composite sous forme analogique est transmis par un canal de communication analogique, puis est converti sous format digital pour être décodé. Les composantes reconstruites provenant du décodeur sont fournies à un système de visualisation stéréoscopique qui reproduit la scène observée de façon à ce que l'information de profondeur qu'elle contient puisse être perçue par un observateur humain. Notre travail porte sur l'encodeur, le décodeur et l'estimation de la région de fixation. Le chapitre 5 décrit en détails ces trois modules.

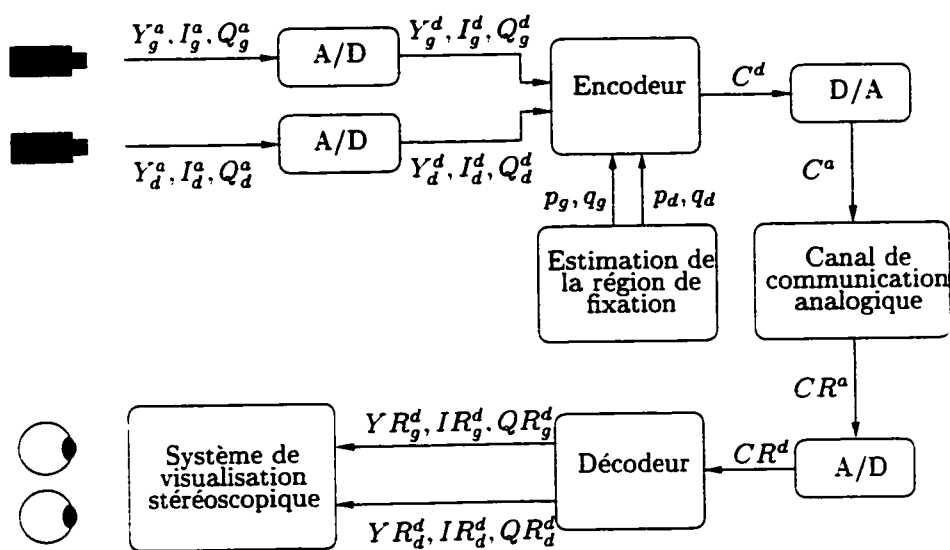


Figure 3.3 : Système de transmission d'images. Les lettres a ou d en exposant indiquent un signal sous forme continue ou digitale et les lettres g et d en indice indiquent les canaux gauche et droit.

Chapitre 4

Les stratégies visuelles pour une tâche de discrimination de profondeur

4.1 Description de l'étude

Les considérations présentées au chapitre 3 nous ont conduit à accomplir notre propre étude psychophysique sur les stratégies visuelles utilisées lors d'une tâche de téléopération. Sur la base de critères présentées dans ce chapitre, une tâche dynamique de discrimination de profondeur a été choisie. Cette tâche consiste à indiquer la direction du mouvement d'objets se déplaçant en profondeur, en présence de distracteurs dynamiques pouvant interférer avec la tâche en masquant temporairement

les objets dont le mouvement doit être identifié. La tâche choisie est typique d'un contexte de téléopération dans la mesure où elle représente une tâche sous-jacente à plusieurs opérations plus complexes, telles que la télémanipulation, le suivi d'objets en mouvement et l'observation d'images dynamiques dans un but d'inspection ou d'interprétation. De plus, cette tâche peut facilement être simulée de façon à ce que l'information de profondeur soit essentiellement apportée par la stéréoscopie. La tâche choisie implique une forte charge perceptive (voir l'annexe B) puisque l'attention doit être soutenue dans la région où les objets se déplacent en profondeur. Une demande d'attention supplémentaire est nécessaire afin de détecter les distracteurs se déplaçant dans le reste de l'image. De plus, l'apparition des distracteurs ainsi que leur mouvement en périphérie sont reconnus comme deux facteurs attracteurs d'attention ou de fixation (voir l'annexe B).

Une situation où un opérateur doit accomplir sa tâche en présence de distracteurs dynamiques a été simulée. Afin d'isoler le système de vision stéréoscopique, les stimuli ont été présentés sous forme de stéréogrammes dynamiques. La tâche principale des sujets consistait à identifier si les objets se déplaçant en profondeur au centre de l'écran approchaient ou s'éloignaient d'eux. Pendant qu'ils accomplissaient cette tâche, des distracteurs se déplaçant en profondeur, soit vers les sujets ou soit dans un plan fronto-parallèle, suivaient des trajectoires allant d'une extrémité à l'autre de l'écran avec 8 directions possibles. Lorsque leur trajectoire croisait la région centrale, les distracteurs pouvaient interférer avec la tâche principale en cachant les objets

se trouvant dans cette région. Les sujets avaient toutefois la possibilité de faire disparaître temporairement les distracteurs de façon à ce qu'ils ne masquent pas la tâche principale dans la région centrale. Pendant qu'ils accomplissaient cette tâche, les mouvements des yeux des sujets étaient enregistrés de façon à ce que des stratégies visuelles puissent être identifiées par la suite.

La présente étude traite de deux aspects fondamentaux de la performance d'un opérateur lors de l'accomplissement de la tâche choisie. Le premier aspect concerne les mouvements des yeux en fonction de la présence de distracteurs. En l'absence de distracteurs, il est prévisible que les fixations de l'opérateur seront concentrées dans la région de l'image associée à la tâche principale. Par contre, il est moins prévisible d'établir si les distracteurs en mouvement induiront des déplacements de fixations à l'extérieur de la région de la tâche principale. Si c'était le cas, toute stratégie d'estimation de la région de fixation devrait impliquer la détection d'objets en mouvement dans une scène et le déplacement de la région de fixation en fonction de certaines caractéristiques de ces objets, tels que leur position, leur type et leur direction de mouvement. D'un autre côté, si les fixations de l'opérateur demeuraient concentrées dans la région de la tâche principale et que les distracteurs étaient tout de même détectés correctement, cela indiquerait que la vision périphérique joue vraisemblablement un rôle important dans l'accomplissement de la tâche. Dans ce cas, la stratégie d'estimation de région de fixation consisterait simplement à choisir la région de fixation la plus grande possible autour du point de fixation de l'observateur.

Le second aspect de la performance de l'opérateur concerne les conditions particulières associées au visionnement stéréoscopique. Lors de l'utilisation de systèmes de présentation stéréoscopique, il est généralement supposé que les sujets placent leur point de fixation sur le plan de l'écran (Hodges et Davis 1993). Il est toutefois possible que les sujets placent plutôt leur plan de fixation en avant ou en arrière de l'écran. Cette situation représente une source potentielle d'information conflictuelle puisque l'accommodation et la vergence deviennent alors dissociées. Elle peut aussi affecter la mesure des points de fixation des sujets puisque les systèmes d'enregistrement des mouvements des yeux dépendent de la même supposition. Ces observations suggèrent de ne pas négliger la géométrie stéréoscopique inhérente à la visualisation des stimuli dans les explications proposées pour tenir compte de certains résultats expérimentaux.

Les conditions expérimentales, les stimuli et la procédure psychophysique sont décrites en détails à l'annexe C. Il est important de mentionner que la présente étude, bien que tout à fait conforme à ce qui se fait dans le domaine de la psychophysique quant à la rigueur du protocole d'expérimentation, ne s'inscrit pas tout à fait dans la cadre de la majorité des études sur le système visuel humain du fait que l'effet de plusieurs facteurs sont mesurés conjointement. Le choix de ne pas isoler et tester systématiquement chaque facteur pouvant influencer les déplacements de fixations a été fait en fonction de notre intérêt pour des conditions représentatives de situations naturelles où plusieurs facteurs sont simultanément présents, et aussi, de façon plus pragmatique, dans le but de limiter l'envergure de notre investigation psychophysique.

En conséquence, notre étude doit être considérée comme un travail préliminaire dont les résultats pourraient éventuellement être utilisés comme point de départ pour une étude dont le but serait la modélisation de certains mécanismes de vision humaine. Cette modélisation n'est toutefois pas notre objectif et notre étude s'avère tout à fait adéquate pour le contexte dans lequel nous désirons exploiter ses résultats, c'est-à-dire pour le développement d'une stratégie d'estimation de région de fixation.

4.2 Résultats

L'écran a été divisé en 21 zones, tel qu'illustré à la figure 4.1. La zone A sous-tend environ 7 deg d'angle visuel, de -3,5 à 3,5 deg. Les zones B à I, J à Q, R à U et S à T s'étendent respectivement de -8,1 à 8,1 deg, -13,7 à 13,7 deg (la hauteur de l'écran) et -18,1 à 18,1 deg (la largeur de l'écran). Les résultats pour différentes conditions sont présentés en indiquant le pourcentage des fixations totales qui ont été mesurées dans chacune des zones.

4.2.1 Direction du mouvement des distracteurs

La première analyse des données a consisté à vérifier l'existence d'une relation entre la direction des distracteurs et la distribution des fixations dans l'image. Pour chaque sujet, les pourcentages de fixations dans chacune des zones de la figure 4.1 ont été calculés pour chaque direction de mouvement des distracteurs. Les données d'un des sujets, représentatives de celles des autres sujets, sont présentés au tableau C.1 de

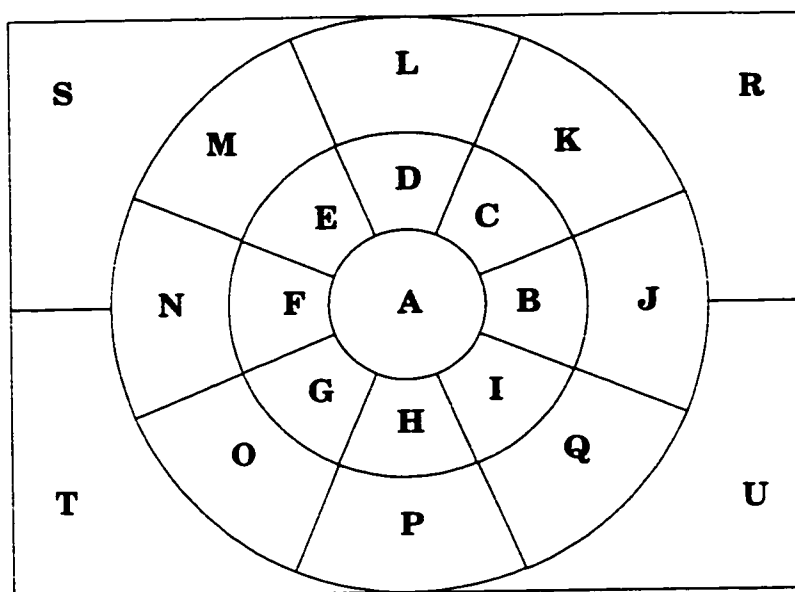


Figure 4.1 : Subdivision de l'écran en 21 zones.

l'annexe C. Ces données montrent qu'il n'existe pas de corrélation significative entre la direction du mouvement des distracteurs et les régions de fixation de l'image. Elles indiquent plutôt que quelques-unes des zones contiennent la majorité des fixations. peu importe la direction du mouvement des distracteurs. La section 4.2.3 décrit plus précisément la distribution des fixations dans ces zones et la section 4.3 propose une explication quant à la cause du type de distribution qui est observé.

4.2.2 Type du mouvement des distracteurs

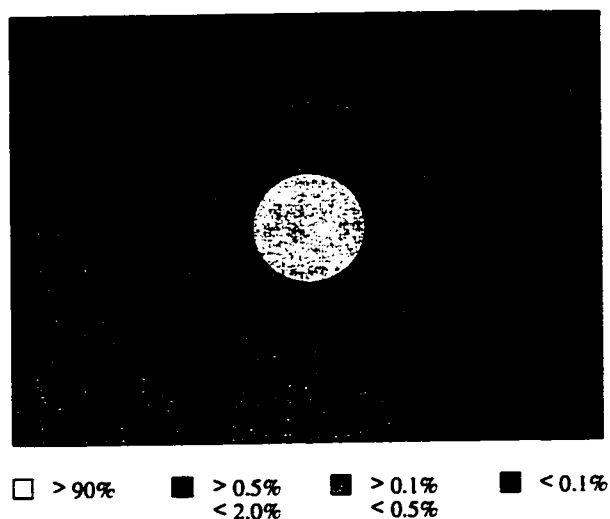
Les données des sujets en fonction du type de mouvement des distracteurs (fronto-parallèle ou vers l'observateur) sont présentées aux tableaux C.2, C.3, C.4, C.5 et C.6 de l'annexe C. Une analyse de ces données indique qu'il n'y a pas de différence significative de la distribution des fixations entre les deux types de mouvement.

4.2.3 Distributions des fixations

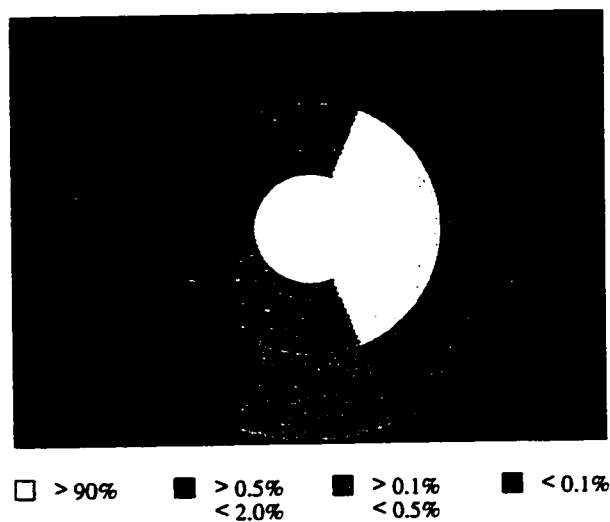
Une analyse des données cumulatives des deux essais de chaque sujet indique que les distributions des fixations peuvent être classées selon deux types principaux de stratégies. En effet, 21 des 24 distributions peuvent être associées à l'un de ces deux types de stratégie. La première stratégie consiste à fixer principalement le centre, *i.e.* plus de 90% des fixations se trouvent dans la zone A. Cette stratégie a été utilisée 12 fois sur 24. La seconde stratégie consiste à fixer une région contenant le centre, mais s'étendant aussi à droite de la région centrale, *i.e.* la somme des pourcentage de fixations dans les zones A,B,C et I excède 90%. Cette stratégie a été utilisée 9 fois sur 24. Un sujet a utilisé une stratégie consistant à jeter des coups d'œil un peu partout (2 essais sur 24) et une autre stratégie a consisté à fixer principalement les zones A et H (1 essai sur 24). Pour ces trois derniers essais, le pourcentage de fixation dans la zone A excède tout de même 60%. Les distributions cumulatives des fixations des sujets pour les deux principales stratégies sont illustrées à la figure 4.2 et les données classées par stratégie sont présentées au tableau C.7 de l'annexe C. La discussion présentée à la section 4.3 propose une explication qui permet de tenir compte des deux principaux types de distribution qui ont été observés.

4.2.4 Effet d'apprentissage

Pour les sujets ayant utilisé la même stratégie lors des deux essais, un effet d'apprentissage se traduisant par une concentration des fixations dans la région centrale



(a)



(b)

Figure 4.2 : Distribution des pourcentages de fixation dans les zones de la figure 4.1 pour les deux principales stratégies visuelles: (a) plus de 90% des fixations se trouvent dans la région centrale (zone A) et (b) plus de 90% des fixations se trouvent dans les zones A, B, C et I.

est observé. Les figures illustrant les données moyennes du premier et du deuxième essai ainsi que la différence entre ces deux essais pour les sujets ayant utilisé la même stratégie lors des deux essais sont présentées au tableau C.8 et aux figures C.3, C.4, C.5, C.6, C.7 et C.8 de l'annexe C.

4.2.5 Calcul du pointage

Des points étaient attribués selon la performance des sujets. Toutefois, la grande variabilité inter-sujet du pointage obtenu ne permet pas de tirer de conclusions reliant le pointage des sujets à leur performance.

4.3 Discussion

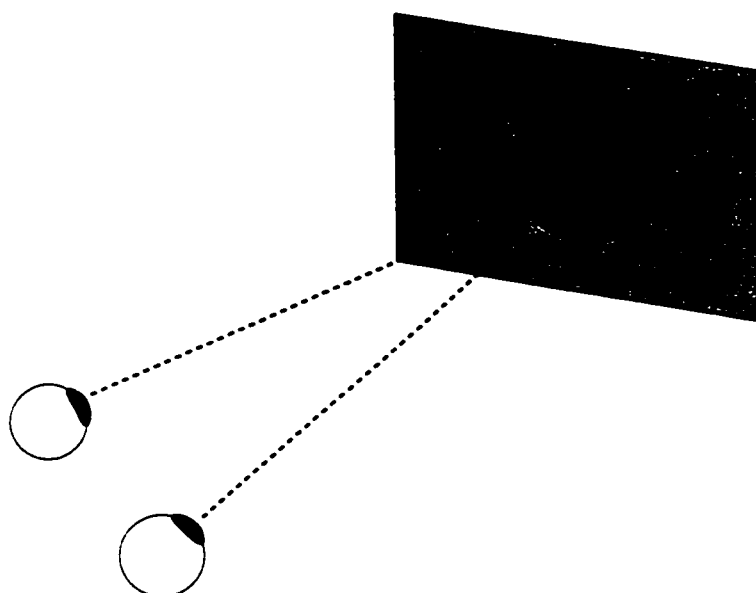
L'étude portait deux aspects de la performance d'un opérateur dans un contexte simulé de téléopération, soit la séquence de mouvements des yeux lors de l'accomplissement de la tâche et l'effet des conditions de visualisation stéréoscopique. En ce qui concerne les séquences de fixations des sujets, deux éventualités avaient été présentées à la section 4.1, soit un déplacement des fixations corrélé avec le mouvement des distracteurs ou soit une absence de déplacement de fixations. Nos résultats supportent la seconde éventualité puisque pour 21 des 24 essais la majorité des fixations mesurées se trouvent à l'intérieur et autour de la région centrale associée à la tâche. Même pour les trois essais n'ayant pas été classifiés selon une des deux stratégies principale, la distribution des fixations dans la région centrale est beaucoup plus élevée que dans les

autres régions. La direction du mouvement et le type de mouvement des distracteurs n'affectent pas significativement la tâche principale. Par contre, les distracteurs sont détectés et traités par les sujets puisque la tâche secondaire est exécutée correctement, mais sans toutefois induire un déplacement de la région de fixation. La vision périphérique, qui est reconnue pour être sensible aux mouvements, est probablement impliquée dans ce processus de détection (Anstis 1986). L'effet d'apprentissage qui est observé favorise aussi la région centrale puisqu'il se traduit par une augmentation des fixations dans cette région, ce qui est en accord avec une stratégie de prédiction favorisant la région centrale. L'observation d'un tel effet d'apprentissage suggère que, dans un contexte de téléopération, une période d'entraînement est nécessaire avant qu'un opérateur puisse accomplir efficacement sa tâche.

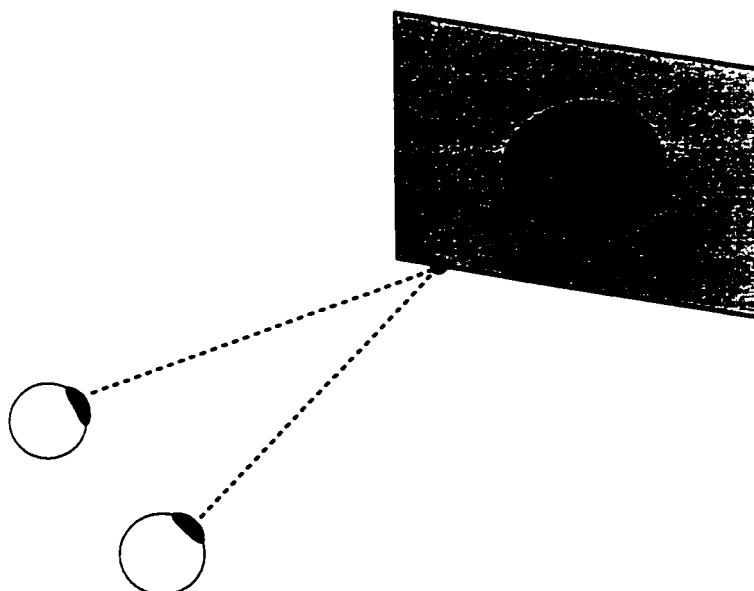
Pour les conditions de visualisation stéréoscopique, nous proposons une explication reliée à la géométrie stéréoscopique afin de tenir compte des fixations se trouvant à droite de la région centrale. Lorsque des écrans stéréoscopiques sont utilisés afin de présenter de l'information visuelle en 3D, il est généralement supposé que les sujets placent leur point de fixation sur l'écran de l'ordinateur (Hodges et Davis 1993) (voir la figure 4.3.a) et conséquemment les valeurs de disparité sont calculées par rapport à cette référence afin de reproduire les objets avec des niveaux de profondeur appropriés. Les systèmes d'enregistrement du mouvement des yeux se basent généralement sur la même hypothèse et mesurent donc non pas un point de fixation dans l'espace tridimensionnel, mais sa projection sur l'écran d'ordinateur. En conséquence, si les

sujets placent leur point de fixation devant l'écran d'ordinateur, les fixations qui sont mesurées sont décalées horizontalement, tel qu'illustré à la figure 4.3.b. Dans la présente expérience, les mouvements de l'œil gauche étaient mesurés. Nous proposons donc que les sujets utilisant la stratégie consistant à fixer la région centrale ainsi que les régions à la droite de celle-ci plaçaient de temps en temps leur point de fixation devant l'écran, ce qui explique la distribution des fixations à la droite de la région centrale (voir la figure 4.2.b). De la même façon, les fixations de l'œil droit devraient se trouver dans la région centrale et dans une région se trouvant à gauche de celle-ci.

L'élaboration d'une stratégie d'estimation de région de fixation pour le codage de séquences stéréoscopiques doit tenir compte des explications proposées pour interpréter les résultats obtenus. La présente étude avait été conçue dans le but de développer une stratégie de prédiction de région de fixation en fonction des déplacements des distracteurs. Implicitement, il était supposé que par des connaissances préalables ou par l'une des méthodes d'estimation présentées au chapitre 2, la position initiale de la région de fixation pourrait être déterminée. À partir de ce point de départ, des séquences de fixations seraient ensuite prédites en fonction des déplacements des distracteurs. Les résultats expérimentaux suggèrent plutôt que, pour les conditions étudiées, une fois la région associée à la tâche principale connue, la meilleure stratégie d'estimation consiste à favoriser cette région plutôt qu'à déplacer les points de fixation en fonction des distracteurs. Cette stratégie d'estimation s'applique à tous les sujets de la présente étude, peu importe la stratégie visuelle



(a)



(b)

Figure 4.3 : Géométrie des systèmes de visualisation stéréoscopique :(a) point de fixation sur l'écran d'ordinateur et (b) point de fixation devant l'écran d'ordinateur.

qu'ils ont utilisées, puisque pour toutes les distributions mesurées, le pourcentage de fixations dans la zone centrale excédait 60%. D'autre part, nos conclusions portant sur la visualisation stéréoscopique suggèrent de favoriser par la suite une zone adjacente à la région correspondant à la tâche principale, située du côté opposé à celui de l'œil qui la regarde. Ces principes seront pris en compte pour l'élaboration des méthodes de codage présentées au chapitre 5.

Chapitre 5

Les méthodes de codage

Les méthodes de codage qui ont été développées compactent l'information stéréoscopique dans l'espace spectral d'un canal vidéo monoculaire. Ces méthodes exploitent la présence d'une région de fixation pour réduire l'information de luminance passe-haut qui est transmise. Une compatibilité avec le standard vidéo NTSC est préservée dans la mesure où la même sous-porteuse et la même région spectrale est utilisée pour l'information de chrominance. Afin de tenir compte de la largeur de bande de différents systèmes de communication, trois méthodes ont été élaborées :

(1) la méthode *studio* lorsque la largeur de bande allouée au signal vidéo est de $2f_{sp} = 7,16 \text{ MHz}$, (2) la méthode *diffusion* lorsque qu'elle est limitée à $f_l = 4,2 \text{ MHz}$, et (3) la méthode *intégrée* qui permet de satisfaire simultanément les contraintes de largeur de bande des deux premières méthodes. La prise en compte des résultats de l'étude sur les mécanismes de fixation du chapitre 4 pour le développement d'une

stratégie d'estimation de région de fixation se fait en deux étapes. Premièrement, de la même façon que dans cette étude, l'hypothèse est faite que la position de la région de l'image correspondant à la tâche principale a été obtenue à l'aide d'une des méthodes d'estimation présentée dans la littérature et rapportée chapitre 2. Cette information, bien qu'elle soit nécessaire pour les deux canaux stéréoscopiques, ne doit être fournie que pour un seul des canaux puisqu'une procédure automatique a été développée pour d'établir la position de la région de fixation dans le second canal stéréoscopique. Cette procédure, qui est représentée par le module d'estimation de la région de fixation à la figure 3.3, est décrite à la section 5.1. Deuxièmement, les résultats de l'étude du chapitre 4 concernant la visualisation stéréoscopique sont pris en compte en favorisant tout d'abord la région correspondant à la tâche principale, puis une région située du côté opposé à celui de l'œil qui la regarde pour chaque canal. L'application de cette stratégie est spécifique à chacune des trois méthodes et est donc présentée avec la description de ces méthodes aux sections 5.2, 5.3 et 5.4. Une résumé comparatif des méthodes conclut le chapitre. L'annexe D contient des détails supplémentaires sur les méthodes de codage qui n'ont pas été inclus dans ce chapitre afin de ne pas alourdir le texte.

5.1 Estimation de la région de fixation

Tel que mentionné au chapitre 4, l'étude qui y est présentée ne fournit pas une solution complète au problème de l'estimation d'une région de fixation. Cette étude

traitait de deux aspects de ce problème lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur : les déplacements de point de fixation en fonction de la présence de distracteurs et les conditions spécifiques de visualisation stéréoscopique. La situation étudiée supposait une connaissance de la région de l'image correspondant à la tâche principale et impliquait de nombreux distracteurs pouvant potentiellement induire des déplacements de fixation. Les résultats de l'étude ont démontré que, pour la tâche étudiée, les distracteurs peuvent être identifiés par les sujets sans toutefois induire de déplacements significatifs de la région de fixation. Ces résultats suggèrent donc une stratégie d'estimation qui consiste simplement à choisir la région de fixation comme étant celle correspondant à la région de l'image associée à la tâche principale. La taille de la région de fixation devrait être la plus grande possible, bien qu'en pratique elle soit limitée par l'espace spectral qui est alloué à l'information de hautes-fréquences par les différentes méthodes de codage, comme il le sera expliqué aux sections suivantes. Le problème d'estimation de la région de fixation est donc réduit à la seule estimation de sa position. Cette position correspond au point de fixation estimé de l'observateur.

Pour le développement des méthodes de codage, la même hypothèse que pour l'étude du chapitre 4 est faite, c'est-à-dire que nous supposons que la position de la région de fixation est connue pour chaque champ d'au moins un des canaux stéréoscopiques et que cette information a été obtenue à l'aide d'une des méthodes d'estimation proposée dans la littérature et présentée au chapitre 2. Comme les

trois méthodes de codage requièrent cette information pour les deux canaux, il est nécessaire d'estimer la position correspondante de la région de fixation dans le second canal stéréoscopique. Cette estimation est possible à l'aide d'une procédure d'appariement par corrélation unidimensionnelle qui n'est appliquée qu'à un bloc A du premier canal positionné autour du centre de la région de fixation. On effectue la corrélation du bloc A , de taille 8×8 , avec des blocs B_i de taille 8×8 du second canal se trouvant sur la même ligne horizontale selon l'équation 5.1 (nous supposons que la disparité verticale est négligeable ; si ce n'est pas le cas, il est toujours possible d'accomplir une rectification des images (voir l'annexe A)). La position du bloc B_i maximisant la mesure de corrélation r de l'équation 5.1 fournit le meilleur estimé de la position de la région de fixation dans le second canal.

$$r = \frac{\sum_{n_1} \sum_{n_2} A(n_1, n_2) B_i(n_1, n_2)}{\sqrt{\sum_{n_1} \sum_{n_2} A^2(n_1, n_2) \sum_{n_1} \sum_{n_2} B_i^2(n_1, n_2)}} \quad (5.1)$$

La grande taille de la région de fixation assure que le point de fixation effectif de l'observateur se trouve dans la région de fixation prédite. Les erreurs de positionnement qui peuvent se produire durant l'appariement n'affectent donc pas significativement les méthodes de codage. Contrairement à la majorité des algorithmes de compression stéréoscopique, la procédure d'appariement n'est appliquée ici qu'à un seul bloc du premier canal.

5.2 Méthode studio : luminance limitée à $2f_{sp} = 7,16 \text{ MHz}$

Deux variations de la méthode *studio* ont été développées afin de compacter l'information stéréoscopique dans l'espace spectral d'un canal vidéo monoculaire (voir la figure 3.2.a et 3.2.c). Le traitement des composantes de luminance est le même pour les deux méthodes, alors que celui des composantes de chrominance diffère par le fait qu'une des deux méthodes implique un filtrage spatio-temporel afin d'exploiter les trous spectraux de Fukinuki (Fukinuki et Hirano 1984), tandis que l'autre ne requiert qu'un filtrage spatial. L'idée principale de la méthode proposée consiste à décomposer l'information de luminance de chaque canal en une composante passe-bas et une composante passe-haut et à recombinaison ces composantes avec celles de chrominance afin d'occuper l'espace spectral de la manière illustrée à la figure 5.1.

5.2.1 Traitement des composantes de luminance

Les composantes de luminance des deux canaux, YG et YD , sont décomposés en composantes passe-bas, $YPBG$ et $YPBD$, et en composantes passe-haut, $YPHG$ et $YPHD$, à l'aide de filtres spatiaux 2D séparables à RIF (réponse impulsionnelle finie) (Oppenheim et Schafer 1989), dont les fréquences de coupure horizontale et verticale correspondent à f_{sp} et $f_v/2$ respectivement. Il est bon de rappeler que $f_{sp} = 3,58 \text{ MHz}$ et $f_v = 120 \text{ c/hi}$. Les composantes de haute-fréquences $YPHG$ et $YPHD$ oc-

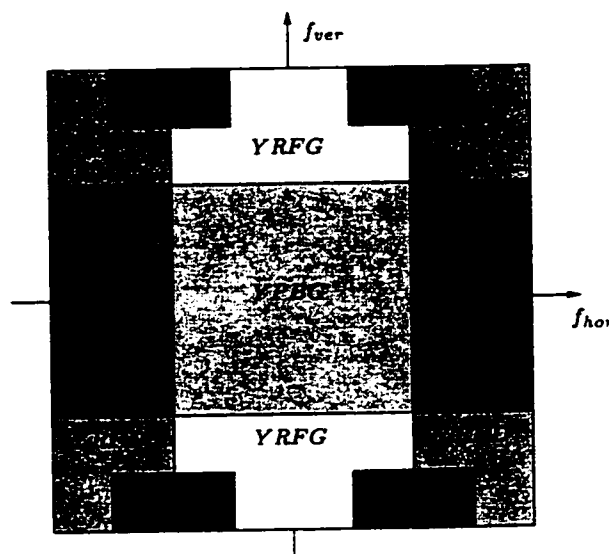


Figure 5.1 : Division de l'espace spectral avec la méthode *studio*. Les symboles $\{YPBG, YPBD\}$, $\{YRFG, YRFD\}$ et *chrom.* représentent respectivement les composantes de luminance passe-bas gauche et droite, de luminance passe-haut gauche et droite associées à la région de fixation et les composantes de chrominance.

cupent un espace spectral nettement supérieur à celui qui est alloué à l'information de hautes-fréquences à la figure 5.1. Par contre, puisque cette information ne doit pas être transmise pour toute l'image, mais seulement pour la région de fixation, l'étendue spectrale des composantes passe-haut peut être réduite en exploitant la propriété de mise à l'échelle de la transformée de Fourier ($x[cn] \leftrightarrow \frac{1}{|c|} X(e^{j\frac{w}{c}})$). Chaque composante de haute-fréquences est donc limitée à une région de fixation dont la largeur et la hauteur correspondent à la moitié de la taille originale des champs. Plutôt que d'utiliser directement la position estimée de la région de fixation pour chaque canal, une position décalée vers la gauche pour le point de vue droit et vers la droite pour le point de vue gauche est plutôt utilisée. De cette façon, une plus grande partie de la

région de fixation se trouve du côté opposé à l'œil qui la regarde pour chaque canal. En conséquence, conformément aux résultats de l'étude sur les processus de fixation du chapitre 4, la probabilité que le point de fixation de l'observateur se trouve dans la région de fixation est augmentée. Le décalage correspond à un peu plus du quart de la largeur de la région de fixation, ce qui est équivalent à environ 2,5 arcdeg pour une distance d'observation de 57 cm, soit la taille de la région adjacente à la région centrale de l'expérience du chapitre 4.

Les composantes de hautes-fréquences *YPHG* et *YPHD* limitées à la région de fixation sont par la suite dilatées par deux dans les directions horizontales et verticales. puis filtrées spatialement à f_{sp} et $f_v/2$ afin d'éliminer les répétitions spectrales. Les composantes de hautes-fréquences ainsi traitées, appelées *YRFG* et *YRFD*, occupent un espace spectral dont la taille correspond maintenant à celle qui leur est allouée à la figure 5.1. Les composantes de luminance passe-bas *YPBG* et *YPBD* et les composantes passe-haut de la région de fixation *YRFG* et *YRFD*, ainsi que celles de chrominance, dont le traitement est décrit un peu plus loin, sont ensuite positionnées dans l'espace spectral tel qu'illustré à la figure 5.1, à l'aide d'opérations de modulation et de filtrage (le filtrage permet d'éliminer les duplications de composantes introduites par la modulation). Le traitement des composantes de luminance est présenté à l'aide de schémas bloc aux figures 5.2 et 5.3 et l'effet de ce traitement dans le domaine spectral pour les composantes passe-haut est représenté à la figure 5.4. Cette figure est importante car elle illustre les principes fondamentaux qui sont exploités par

toutes les méthodes de codage présentées dans ce chapitre. La figure 5.4.a illustre la composante passe-haut obtenue en soustrayant de la composante de luminance originale la composante de luminance passe-bas :

$$Y_{PH} = Y - Y_{PB}. \quad (5.2)$$

La composante passe-haut Y_{PH} est limitée spatialement à une région de fixation de taille $L/2$ par $H/2$, où L et H représentent respectivement la largeur et la hauteur d'un champ. Cette composante est par la suite agrandie (zoom) par un facteur de 2 dans les directions horizontales et verticales :

$$Y_{HP}'[m, n] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} Y_{HP}[j, k] \delta[m - 2j, n - 2k] \quad (5.3)$$

Le spectre de $Y_{HP}'[m, n]$, $Y_{HP}'(e^{j\omega_x}, e^{j\omega_y})$, est égal à $Y_{HP}(e^{2j\omega_x}, e^{2j\omega_y})$. La transformée de Fourier du signal agrandi est donc une version mise à l'échelle de la transformée de Fourier du signal original qui se répète périodiquement, tel qu'illustré à la figure 5.4.b. Les répétitions spectrales sont éliminées avec un filtrage passe-pas dont les fréquences de coupure horizontale et verticale sont respectivement de f_{sp} (le signal est échantillonné à $2f_{sp}$) et $f_v/2$ (voir la figure 5.4.c). Après cette étape, les composantes passe-haut occupent un espace spectral réduit et il ne reste qu'à les positionner dans l'espace alloué au signal vidéo composite. Ce positionnement se fait à l'aide d'une opération de modulation sur le signal vidéo 1D, tel qu'illustré à

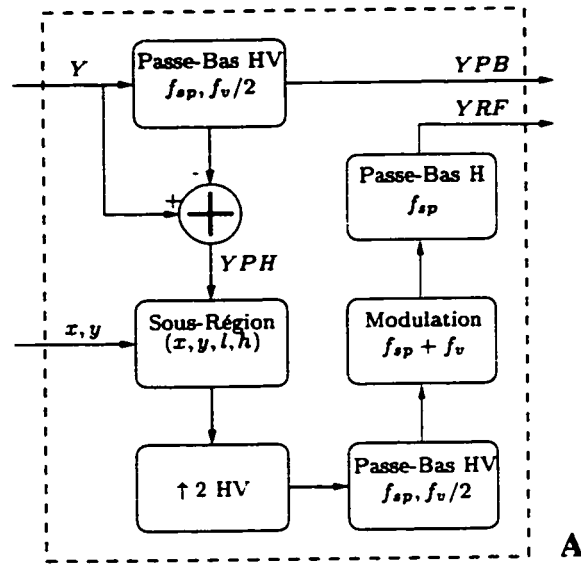


Figure 5.2 : Décomposition de l'information de luminance en composantes passe-bas et passe-haut avec la méthode *studio*. Y , YPH , YPB et YRF représentent respectivement les composantes de luminance originale, passe-haut, passe-bas et passe-haut associée à la région de fixation. x et y représentent la position décalée de la région de fixation et l et h la largeur et la hauteur de la région de fixation, avec $l = L/2$ et $h = H/2$, la moitié de la largeur et de la hauteur d'un champ.

la figure 5.4.d. Cette modulation implique toutefois une duplication de composante ($x \cos(\omega_0 n) \leftrightarrow \pi \sum_{k=-\infty}^{\infty} [X(\omega - \omega_0 + 2\pi k) + X(\omega + \omega_0 + 2\pi k)]$). Cette duplication est éliminée avec un filtrage passe-bas horizontal à f_{sp} , tel qu'illustré à la figure 5.4.e. Après cette opération, la composante passe-haut gauche occupe la position qui lui est assignée pour la formation du signal vidéo composite. Une opération de modulation additionnelle est accomplie pour positionner la composante passe-haut droite, tel qu'illustré à la figure 5.4.f. Cette fois, les duplications de composantes n'ont pas à être éliminées car elles se produisent à une fréquence de $2f_{sp}$.

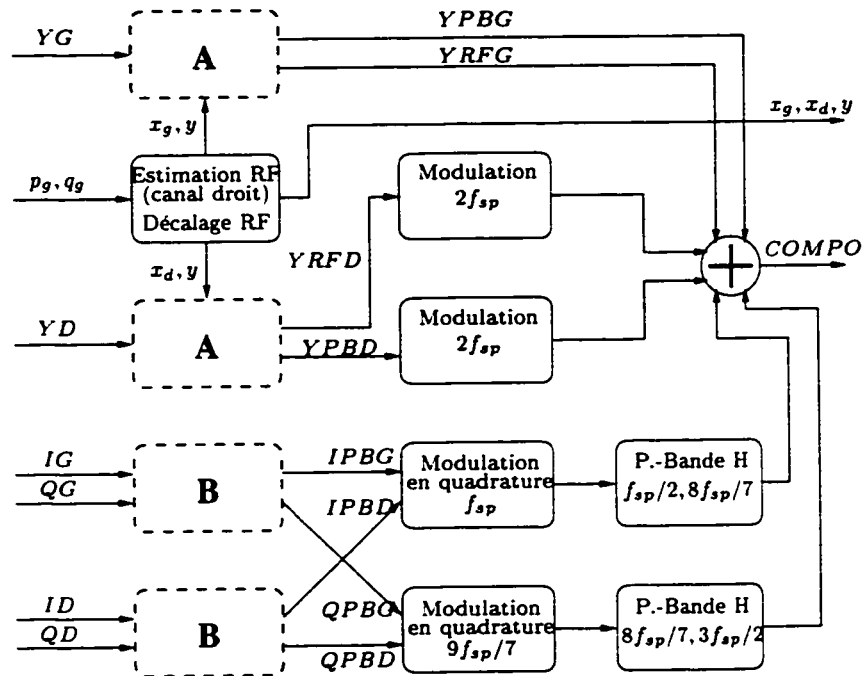


Figure 5.3 : Formation du signal vidéo composite avec la méthode *studio*. $\{YG, YD\}$ et $\{IG, QG, ID, QD\}$ correspondent aux composantes originales de luminance et de chrominance des canaux gauche et droit et $COMPO$ au signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation dans le canal gauche, alors que x_g et x_d représentent la position décalée de la région de fixation dans les canaux gauche et droit, avec $y = q_g$ pour les deux canaux. Le bloc **A** se trouve à la figure 5.2 et le bloc **B** à la figure 5.5.

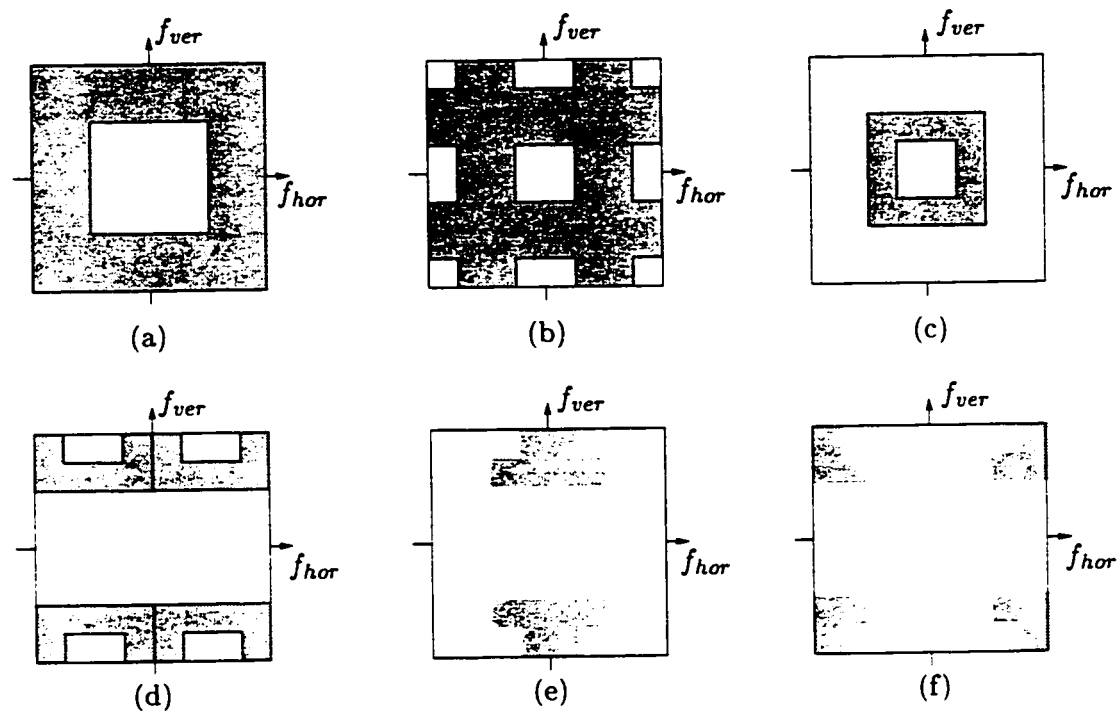


Figure 5.4 : Traitement des composantes passe-haut dans le domaine spectral avec la méthode *studio*: (a) composante passe-haut originale, (b) agrandissement (zoom) de la région de fixation, (c) élimination des répétitions spectrales, (d) positionnement spectral par modulation, (e) élimination de la duplication de composante et (f) modulation additionnelle pour la composante droite.

5.2.2 Traitement spatial des composantes de chrominance

L'espace spectral disponible pour les composantes de chrominance de la figure 5.1 s'étend horizontalement de $f_{sp} - f_{sp}/2$ à $f_{sp} + f_{sp}/2$ et verticalement de $f_v - f_v/4$ à f_v . Avec le standard NTSC, les composantes I et Q , limitées respectivement à $3f_{sp}/7 = 1,5$ MHz et $f_{sp}/7 = 0,5$ MHz, subissent une modulation d'amplitude en quadrature avec bande latérale résiduelle (VSB) (Gibson 1989) de façon à ce que leur occupation spectrale n'excède pas $f_l = 4,2$ MHz. Avec la méthode *studio*, un espace spectral additionnel est disponible pour les composantes de chrominance, mais quatre composante doivent être insérées dans cet espace. La première approche qui est proposée consiste à moduler en quadrature avec bande latérale résiduelle (VSB) l'amplitude des deux composantes I pour occuper l'espace spectral entre $f_{sp}/2$ et f_l et à moduler en quadrature les deux composantes Q pour occuper l'espace spectral entre f_l et $3f_{sp}/2$. Le traitement des composantes de chrominance selon cette approche est illustré à l'aide de schémas bloc aux figures 5.3 et 5.5 et leur occupation spectrale est présentée à la figure 5.6.

5.2.3 Traitement spatio-temporel des composantes de chrominance

Des améliorations à la première méthode de traitement des composantes de chrominance peuvent être apportées en exploitant plus efficacement l'espace spectral du plan fréquentiel vertical-temporel. Les régions appelées *trous spectraux de Fukinuki*

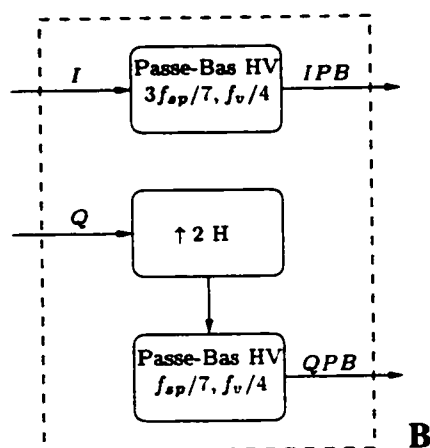


Figure 5.5 : Traitement des composantes de chrominance avec la méthode *studio*. I , Q , IPB et QPB représentent respectivement les deux composantes de chrominance originales et passe-bas.

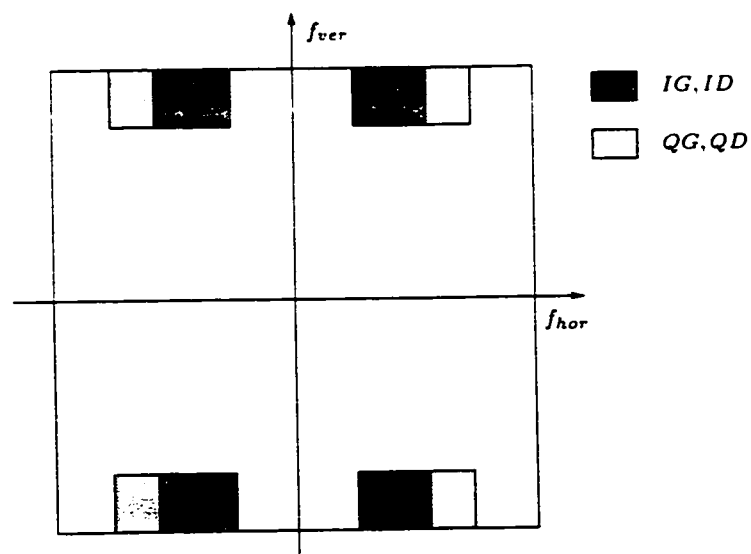


Figure 5.6 : Occupation spectrale des composantes de chrominance avec la méthode *studio*. L'amplitude des composantes $IPBG$ et $IPBD$ est modulée en quadrature avec bande latérale résiduelle et celle des composantes $QPBG$ et $QPBD$ est modulée en quadrature.

(Fukinuki et Hirano 1984) dans les premier et troisième quadrants de la figure 3.2.c ne contiennent pas d'information de chrominance dans le standard NTSC. La seconde approche proposée consiste à placer l'information de chrominance d'un des deux canaux dans ces *trous*. La sous-porteuse de chrominance f_{sp} du standard NTSC est utilisée pour moduler en quadrature l'amplitude des composantes I et Q d'un canal stéréoscopique. Une seconde sous-porteuse, f'_{sp} , dont la phase est inversée par rapport à f_{sp} à tous les deux champs, module en quadrature l'amplitude des composantes I et Q de l'autre canal. Les relations de phase des deux sous-porteuses est illustrée aux figures D.5 et D.6 de l'annexe D. Cette approche implique un filtrage temporel au récepteur afin de récupérer les quatre composantes de chrominance. Idéalement, un filtrage temporel devrait aussi être accompli à l'émetteur afin d'éliminer les interférences entre les composantes de luminance et de chrominance, mais en pratique un tel filtrage ne s'est pas avéré nécessaire. Le traitement des composantes de chrominance selon cette approche est illustré à l'aide de schémas bloc aux figures 5.7 et 5.8 et leur occupation spectrale est présentée à la figure 5.9.

5.2.4 Reconstruction des composantes

Au récepteur, pour la reconstruction des images à partir du signal vidéo composite, l'information de chrominance est d'abord séparée de celle de luminance à l'aide d'un filtre passe-bande horizontal dont le bande passante s'étend de $f_{sp}/2$ à $3f_{sp}/2$ et d'un filtre passe-haut vertical dont la fréquence de coupure est à $3f_v/4$

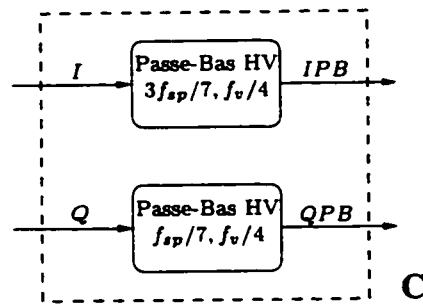


Figure 5.7 : Traitement des composantes de chrominance avec la méthode *studio* avec trous de Fukinuki. I , Q , IPB et QPB représentent respectivement les deux composantes de chrominance originales et passe-bas.

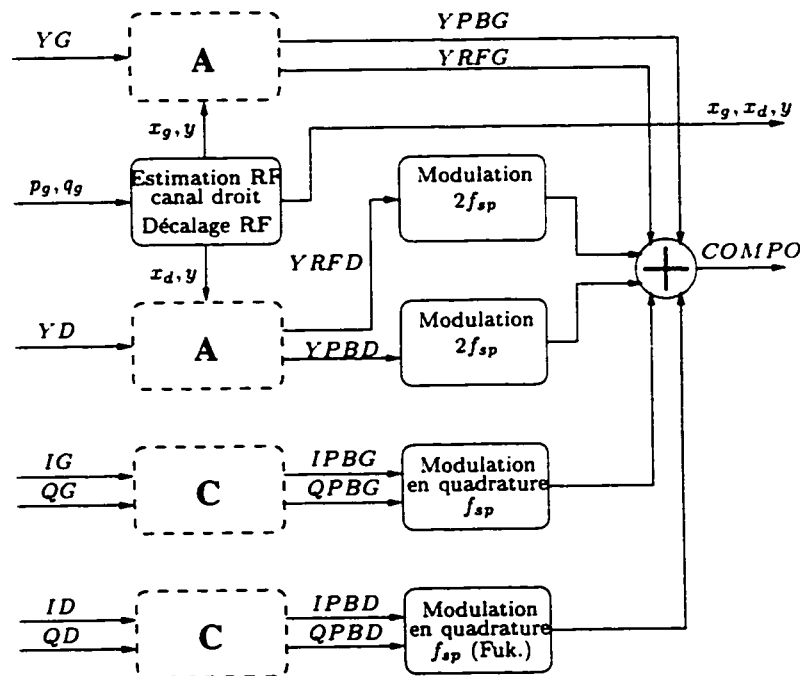


Figure 5.8 : Formation du signal vidéo composite avec la méthode *studio* avec trous de Fukinuki. $\{YG, YD\}$ et $\{IG, QG, ID, QD\}$ correspondent aux composantes originales de luminance et de chrominance des canaux gauche et droit et $COMPO$ au signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation pour le canal gauche, alors que x_g et x_d représentent la position décalée de la région de fixation dans les canaux gauche et droit, avec $y = q_g$ pour les deux canaux. Le bloc A se trouve à la figure 5.2 et le bloc C à la figure 5.7.

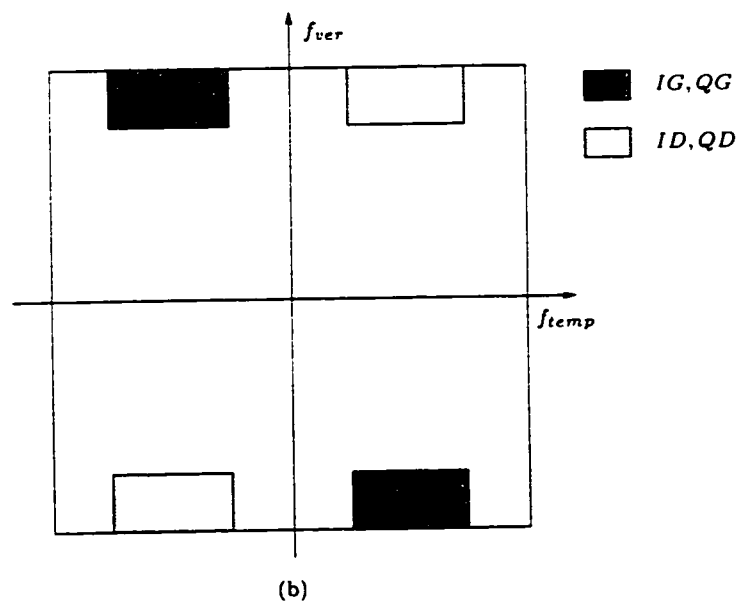
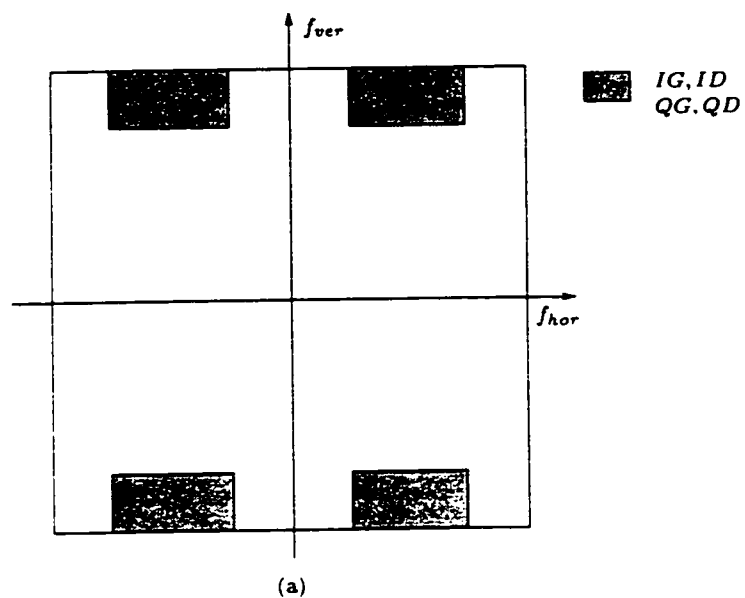


Figure 5.9 : Occupation spectrale des composantes de chrominance avec la méthode *studio* avec trous de Fukinuki. Les composantes de chrominance des canaux gauche et droit (a) occupent le même espace dans le plan fréquentiel horizontal vertical, mais (b) sont séparées dans le plan temporel-vertical.

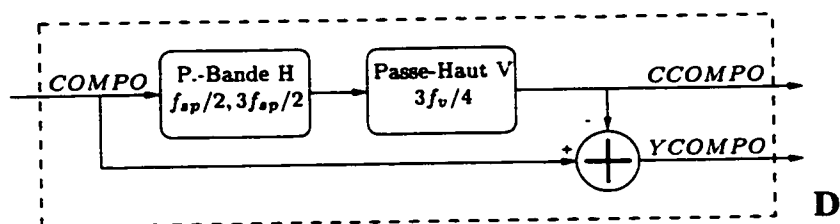


Figure 5.10 : Séparation des composantes de luminance et de de chrominance à partir du signal vidéo composite. *COMPO* représente le signal vidéo composite et *CCOMPO* et *YCOMPO* les parties de ce signal contenant respectivement l'information de chrominance et de luminance.

(voir la figure 5.10). Par la suite, les composantes de luminance sont récupérées à l'aide d'opérations de filtrage, de démodulation et de décimation (voir les figures 5.11 et 5.12). Afin de reconstruire les composantes de luminance, la position de la région de fixation dans les deux canaux stéréoscopiques doit être transmise. Nous supposons que cette information est encodée dans la portion du signal vidéo contenant les lignes inactives. Pour la méthode avec trous de Fukinuki, un filtrage temporel sur 3 champs est accompli afin de récupérer les composantes de chrominance, alors que pour la méthode sans trous de Fukinuki, seul un filtrage spatial est nécessaire. La reconstruction des composantes transmises selon les méthodes *studio* et *studio* avec trous de Fukinuki est illustrée aux figures 5.13 et 5.14 respectivement.

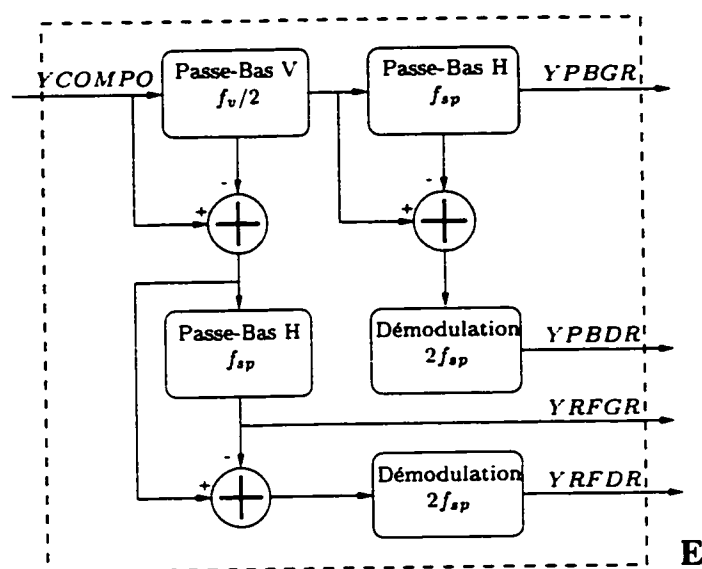


Figure 5.11 : Séparation des composantes de luminance avec la méthode *studio*. $YCOMPO$ représente la partie du signal vidéo composite contenant l'information de luminance et $YPBGR_i$, $YPBDR_i$, $YRFGRI$ et $YRFDR_i$ les composantes de luminance reconstruites passe-bas et passe-haut de la région de fixation pour les canaux gauche et droit.

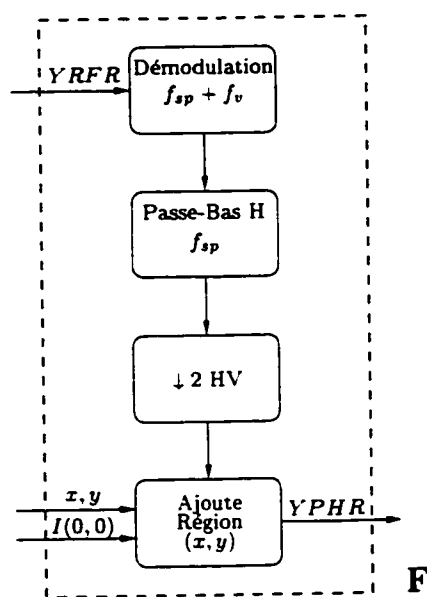


Figure 5.12 : Reconstruction des composantes de luminance passe-haut avec la méthode *studio*. $YRFR$ et $YPHR$ correspondent respectivement aux composantes de luminance reconstruites passe-haut associées à la région de fixation et à l'image entière. $I(0, 0)$ représente une image de la taille d'un champ dont tous les éléments sont égaux à 0 et x et y correspondent à la position de la région de fixation.

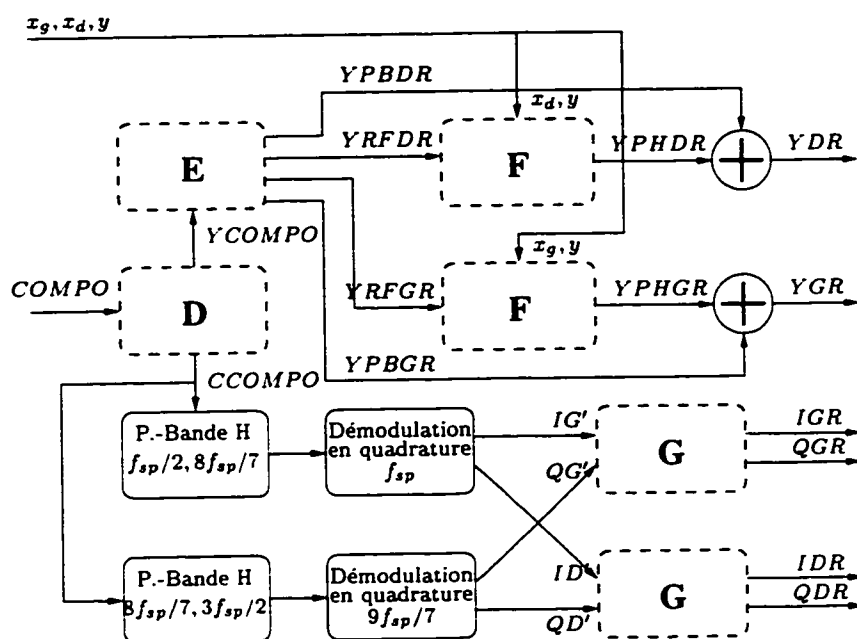


Figure 5.13 : Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode *studio*. *COMPO* correspond au signal vidéo composite et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g, x_d représentent la position horizontale de la région de fixation dans les canaux gauche et droit, et y la position verticale de cette région dans les deux canaux. Le bloc **D** se trouve à la figure 5.10, le bloc **E** à la figure 5.11, le bloc **F** à la figure 5.12 et le bloc **G** à la figure 5.15.

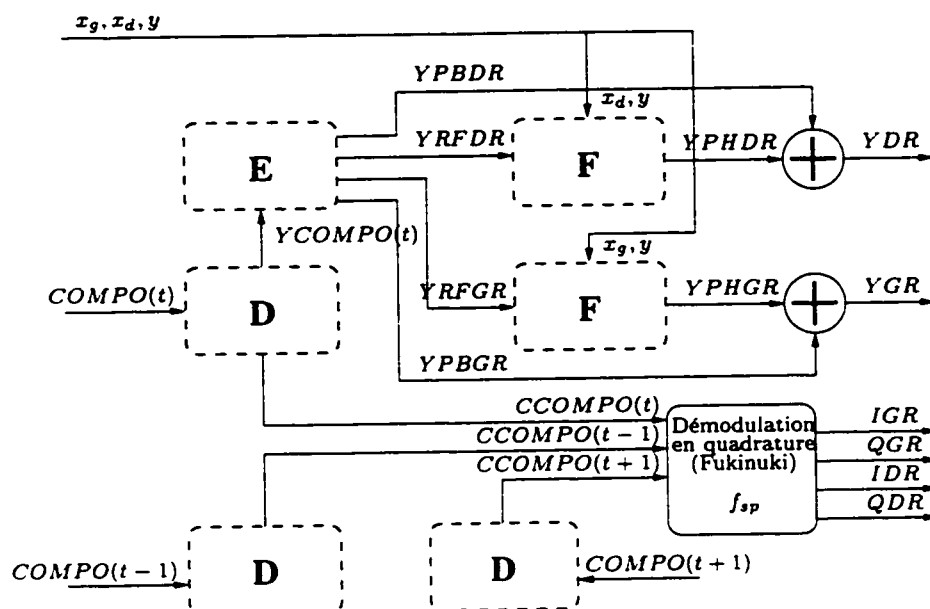


Figure 5.14 : Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode *studio* avec trous de Fukinuki. $COMPO(t-1)$, $COMPO(t)$ et $COMPO(t+1)$ représentent respectivement le signal vidéo composite à l'instant précédent, l'instant courant et l'instant suivant. De même $CCOMPO(t-1)$, $CCOMPO(t)$ et $CCOMPO(t+1)$ correspondent à la partie du signal vidéo contenant l'information de chrominance à l'instant précédent, l'instant courant et l'instant suivant. $YCOMPO(t)$ représente la partie du signal vidéo contenant l'information de luminance à l'instant courant et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ correspondent aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g et x_d représentent la position horizontale de la région de fixation dans les canaux gauche et droit, et y la position verticale de cette région dans les deux canaux. Le bloc **D** se trouve à la figure 5.10, le bloc **E** à la figure 5.11 et le bloc **F** à la figure 5.12.

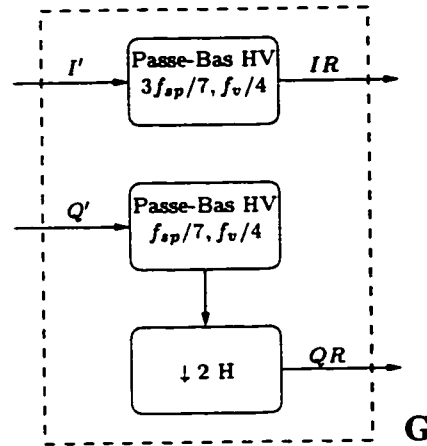


Figure 5.15 : Reconstruction des composantes de chrominance avec la méthode *studio*. I' , Q' , IR et QR correspondent respectivement aux deux composantes de chrominance démodulées et reconstruites.

5.3 Méthode diffusion : luminance limitée à $f_l = 4,2$ MHz

Les méthodes de compression de l'approche *diffusion* suivent les mêmes principes que ceux de la méthode *studio*, la principale différence entre les deux approches étant que l'espace spectral horizontal est maintenant limité à $f_l = 4,2$ MHz. En conséquence, la taille horizontale de la région de fixation est réduite à un cinquième de la largeur des champs, ce qui n'est pas assez grand pour profiter du décalage de points de vue suggéré par les résultats de l'étude de mécanismes de fixation du chapitre 4. La position estimée de la région de fixation est donc directement utilisée avec la méthode *diffusion*. Une méthode de codage symétrique et une méthode asymétrique ont été étudiées pour le traitement des composantes de luminance. La méthode asymétrique

a été développée avec comme objectif la prise en compte explicite de la redondance stéréoscopique entre les deux canaux. Cette méthode, qui est décrite en détails à l'annexe D, s'avère toutefois moins efficace que la méthode symétrique. Les méthodes symétrique et asymétrique pour le traitement de la luminance peuvent être combinées à deux méthodes pour le traitement des composantes de chrominance, donnant ainsi lieu à quatre variations de la méthode *diffusion*.

5.3.1 Traitement des composantes de luminance

Les composantes de luminance des deux canaux sont décomposées en composantes passe-bas, $YPBG$ et $YPBD$, et en composantes passe-haut, $YPHG$ et $YPHD$, à l'aide de filtres spatiaux 2D séparables à RIF dont les fréquences de coupure horizontale et verticale sont respectivement $f_l/2$ et $f_v/2$. Les composantes de hautes-fréquences de chaque canal sont ensuite divisées en deux sous-composantes : les hautes fréquences verticales dont les fréquences horizontales sont plus petites que $f_l/2$, $YPH1G$ et $YPH1D$, et les hautes-fréquences horizontales plus élevées que $f_l/2$, $YPH2G$ et $YPH2D$. Ces composantes sont limitées spatialement à la région de fixation, agrandies horizontalement par un facteur de 5 et verticalement par un facteur de 2, puis filtrées passe-bas afin d'éliminer les répétitions spectrales. Les composantes de luminance ainsi traitées, appelées $YRF1G$, $YRF1D$, $YRF2G$ et $YRF2D$, sont combinées avec les composantes de luminance passe-bas, $YPBG$ et $YPBD$, et les composantes de chrominance, dont le traitement est décrit un peu plus loin, par des

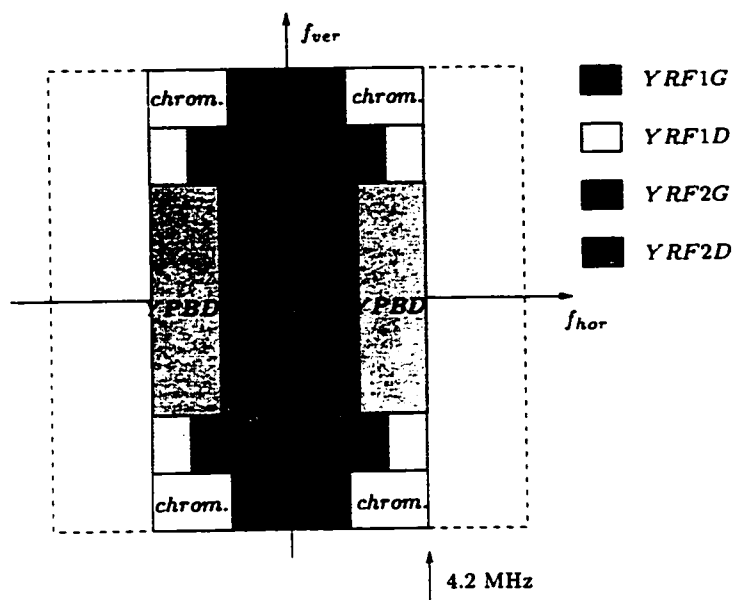


Figure 5.16 : Division de l'espace spectral avec la méthode *diffusion*. $YPBG$, $YPBD$, $YRF1G$, $YRF1D$, $YRF2G$, $YRF2D$ et $chrom.$ représentent respectivement les composantes de luminance passe-bas gauche et droite, les sous-composantes passe-haut gauche et droite associées à la région de fixation et les composantes de chrominance.

opérations de modulation et de filtrage de façon à occuper l'espace spectral tel qu'illustré à la figure 5.16. Le traitement des composantes de luminance est présenté, à l'aide de schémas bloc, aux figures 5.17, 5.18, 5.19 et 5.20 et l'effet de ce traitement dans le domaine spectral est illustré pour les composantes passe-haut à la figure 5.21. L'explication de cette figure repose sur les mêmes principes que ceux de la méthode *studio*, illustrés à la figure 5.4, à la différence que l'information passe-haut de chaque canal est divisée en deux sous-composantes.

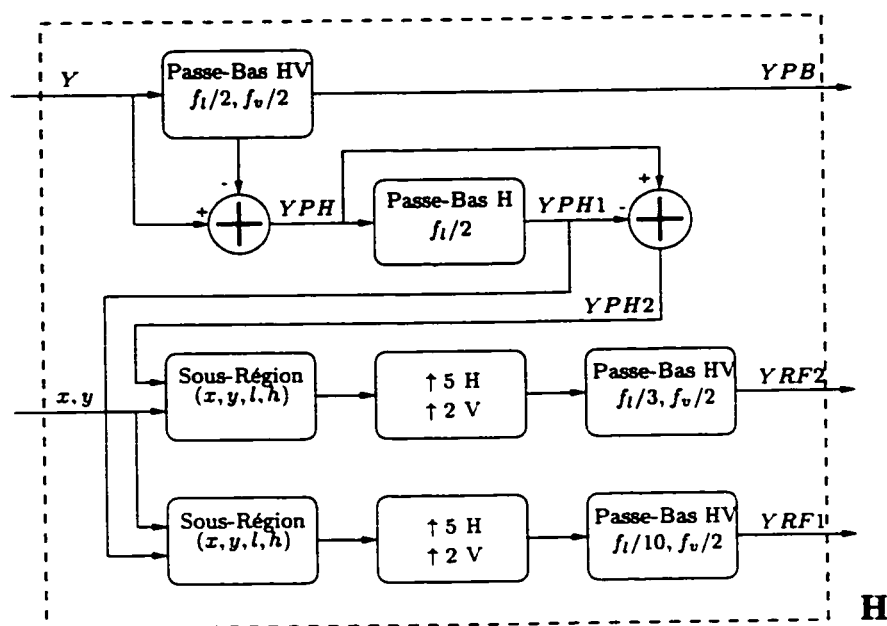


Figure 5.17 : Décomposition de l'information de luminance en composantes passe-bas et passe-haut avec la méthode *diffusion*. Y , YPB , $YRF1$ et $YRF2$ correspondent aux composantes de luminance originale, passe-bas et aux deux sous-composantes de luminance passe-haut associées à la région de fixation. x et y représentent la position horizontale et verticale de la région de fixation et l et h la largeur et la hauteur de la région de fixation, avec $l = L/5$ et $h = H/2$, le cinquième de la largeur d'un champ et la moitié de sa hauteur.

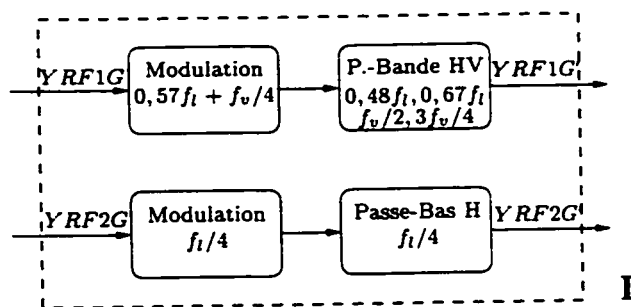


Figure 5.18 : Traitement des composantes passe-haut du canal gauche avec la méthode *diffusion*. $\{YRF1G, YRF2G\}$, $\{YRF1G', YRF2G'\}$ représentent les deux sous-composantes de luminance passe-haut associées à la région de fixation et ces sous-composantes modulées.

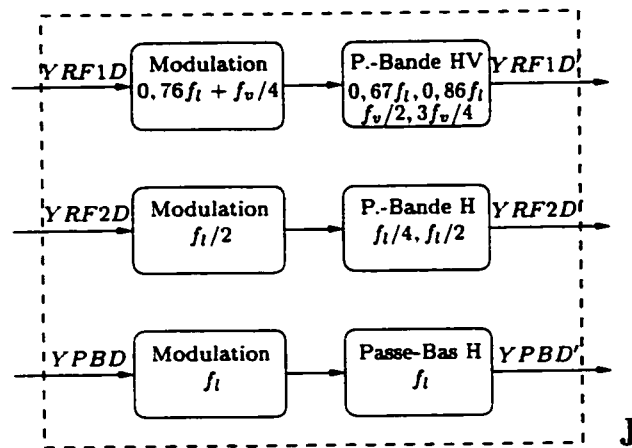


Figure 5.19 : Traitement des composantes passe-haut du canal droit avec la méthode *diffusion*. $YPBD$, $\{YRF1D, YRF2D\}$, $YPBD'$, $\{YRF1D', YRF2D'\}$ représentent la composante de luminance passe-bas, les deux sous-composantes de luminance passe-haut associées à la région de fixation et ces trois composantes modulées.

5.3.2 Traitement des composantes de chrominance

Les méthodes symétrique et asymétrique (voir l'annexe D) pour le traitement des composantes de luminance peuvent être combinées à deux méthodes de traitement des composantes de chrominance. La première méthode consiste à exploiter les *trous* de Fukinuki (Fukinuki et Hirano 1984) pour transmettre les composantes de chrominance d'un canal en modulant en quadrature leur amplitude avec bande latérale résiduelle (VSB). Avec cette méthode, un filtrage temporel doit être accompli au récepteur, et idéalement aussi à l'émetteur. La formation du signal vidéo composite selon cette méthode est illustré à l'aide du schéma bloc de la figure 5.22, alors que le traitement des composantes de chrominance est illustré à la figure 5.5.

Lorsqu'il n'est pas possible d'accomplir un filtrage temporel, la largeur de bande

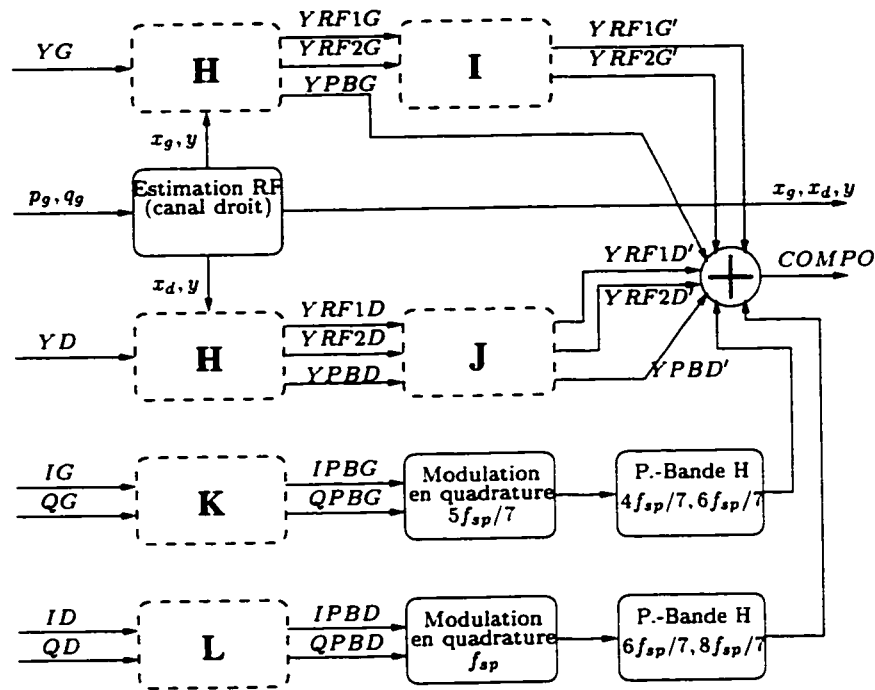


Figure 5.20 : Formation du signal vidéo composite avec la méthode *diffusion*. $\{Y_G, Y_D\}$ et $\{IG, QG, ID, QD\}$ correspondent aux composantes originales de luminance et de chrominance des canaux gauche et droit et *COMPO* au signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation pour le canal gauche. Pour la méthode *diffusion* $x_g = p_g$, x_d représente la position de la région de fixation dans le canal droit et $y = q_g$ pour les deux canaux. Le bloc **H** se trouve à la figure 5.17, le bloc **I** à la figure 5.18, le bloc **J** à la figure 5.19, le bloc **K** à la figure 5.24 et le bloc **L** à la figure 5.25.

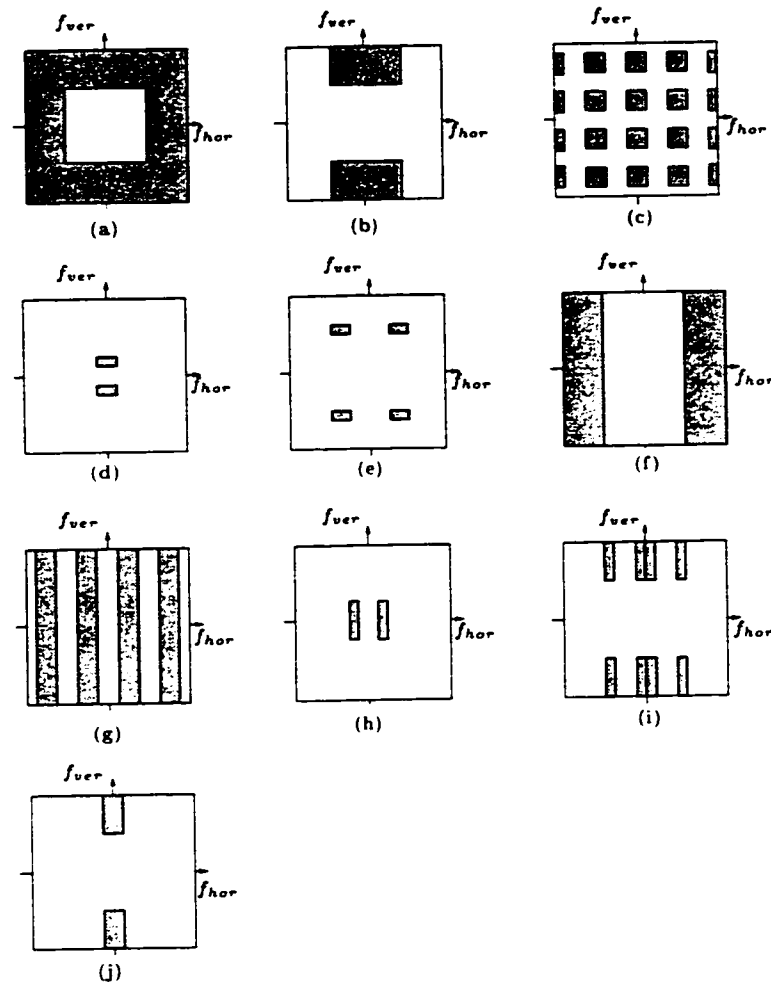


Figure 5.21 : Traitement des composantes passe-haut dans le domaine spectral avec la méthode *diffusion*: (a) composante passe-haut originale, (b) sous-composante passe-haut dont les fréquences horizontales sont plus petites que $f_l/2$, (c) agrandissement de la région de fixation (zoom), (d) élimination des répétitions spectrales, (e) positionnement spectral par modulation, (f) sous-composante passe-haut dont les fréquences horizontales sont plus grandes que $f_l/2$, obtenue en filtrant le signal en (a), (g) agrandissement de la région de fixation (zoom), (h) élimination des répétitions spectrales, (i) positionnement spectral par modulation, (j) élimination de la duplication de composante. Cette figure illustre le traitement de la composante passe-haut du canal gauche ; les mêmes principes s'appliquent pour le traitement de la composante droite.

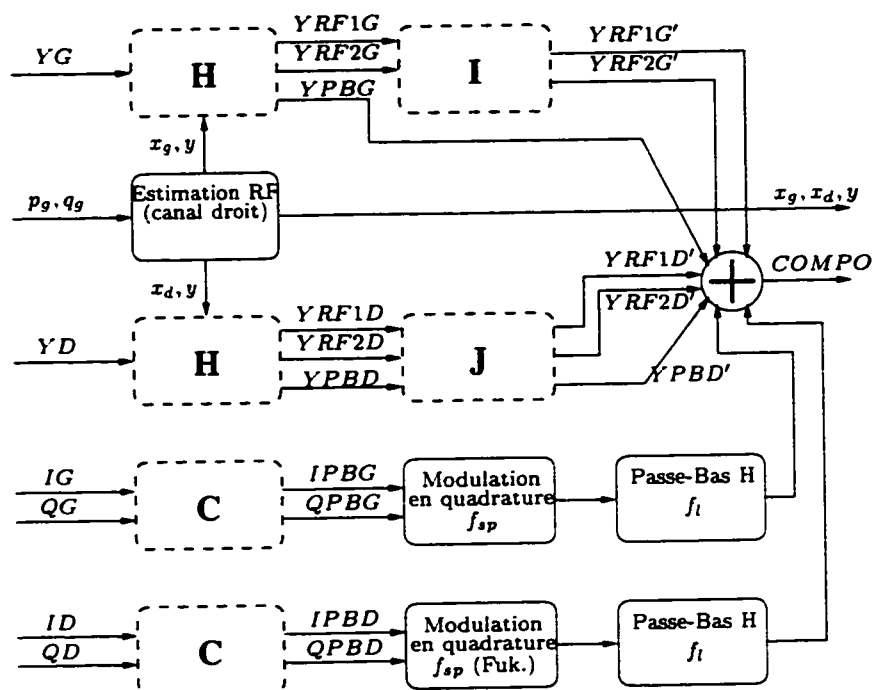


Figure 5.22 : Formation du signal vidéo composite avec la méthode *diffusion* avec trous de Fukinuki. $\{YG, YD\}$ et $\{IG, QG, ID, QD\}$ correspondent aux composantes originales de luminance et de chrominance des canaux gauche et droit et *COMPO* au signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation pour le canal gauche. Pour la méthode *diffusion* $x_g = p_g$, x_d représente la position de la région de fixation dans le canal droit et $y = q_g$ pour les deux canaux. Le bloc **H** se trouve à la figure 5.17, le bloc **I** à la figure 5.18, le bloc **J** à la figure 5.19 et le bloc **C** à la figure 5.5

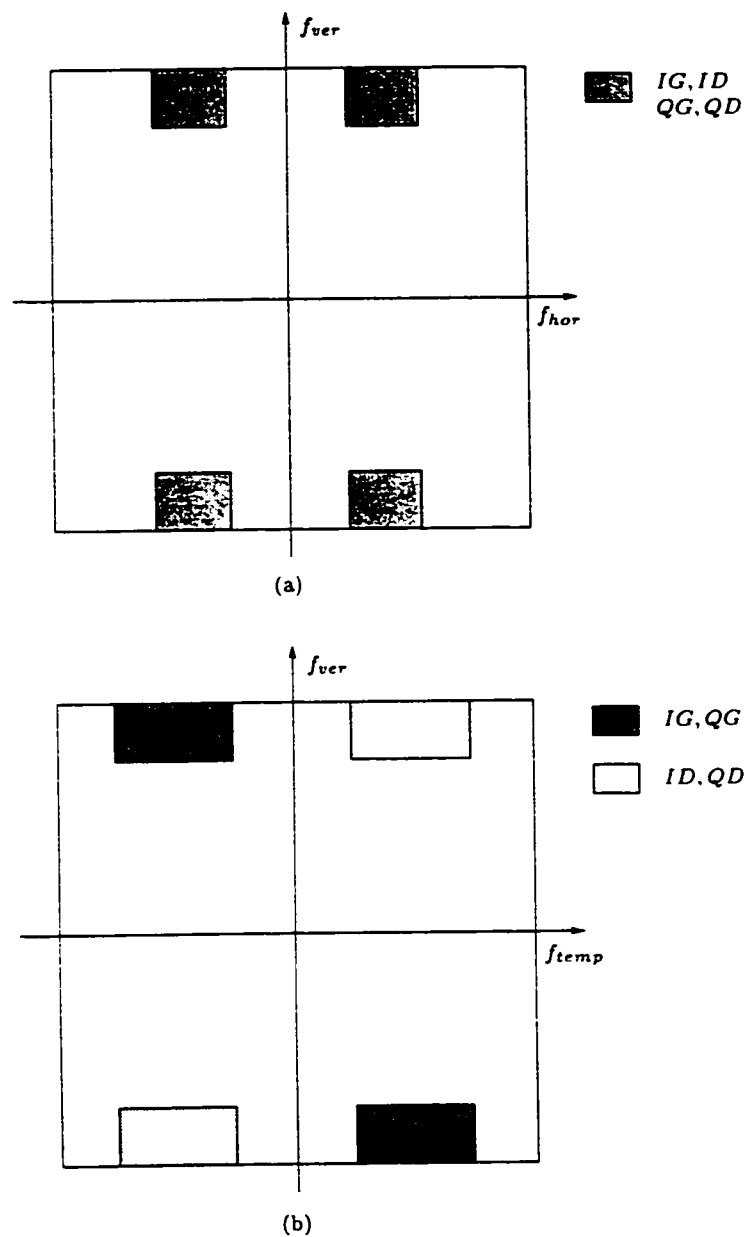


Figure 5.23 : Occupation spectrale des composantes de chrominance avec la méthode *diffusion* avec trous de Fukinuki. Les composantes de chrominance des canaux gauche et droit (a) occupent le même espace dans le plan fréquentiel horizontal-vertical, mais (b) sont séparées dans le plan temporel-vertical.

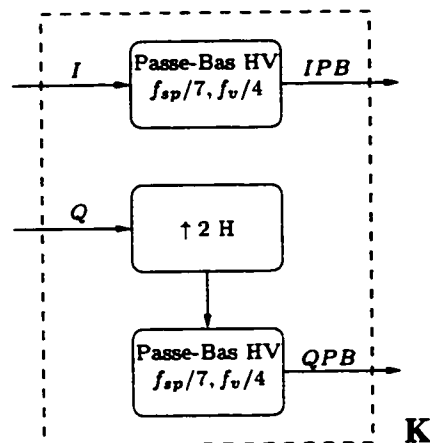


Figure 5.24 : Traitement des composantes de chrominance du canal gauche avec la méthode *diffusion*. I , Q , IPB et QPB représentent respectivement les deux composantes de chrominance originales et passe-bas.

des composantes I peut être réduite à $f_{sp}/7 = 0,5$ MHz. L'amplitude des composantes de chrominance d'un canal est modulée en quadrature pour occuper l'espace spectral réservé à l'information de chrominance allant de $6f_{sp}/7$ à $8f_{sp}/7$. Une seconde sous-porteuse de chrominance $f''_{sp} = 5f_{sp}/7$ est utilisée pour moduler en quadrature l'amplitude des composantes de chrominance du second canal de façon à ce qu'elles occupent l'espace spectral de chrominance allant de $4f_{sp}/7$ à $6f_{sp}/7$. Le traitement des composantes de chrominance avec cette deuxième méthode est illustré aux figures 5.20, 5.24, et 5.25 et leur occupation spectrale est présentée à la figure 5.26.

5.3.3 Reconstruction des composantes

La reconstruction des images au récepteur se fait selon les mêmes principes que ceux de la méthode *studio*. Les schémas bloc illustrant de façon détaillée cette recon-

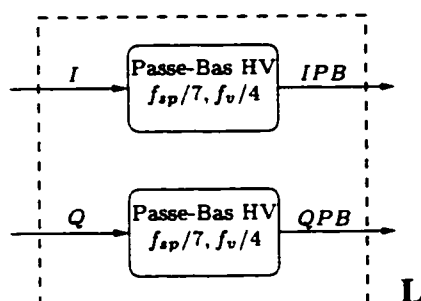


Figure 5.25 : Traitement des composantes de chrominance du canal droit avec la méthode *diffusion*. I , Q , IPB et QPB représentent respectivement les deux composantes de chrominance originales et passe-bas.

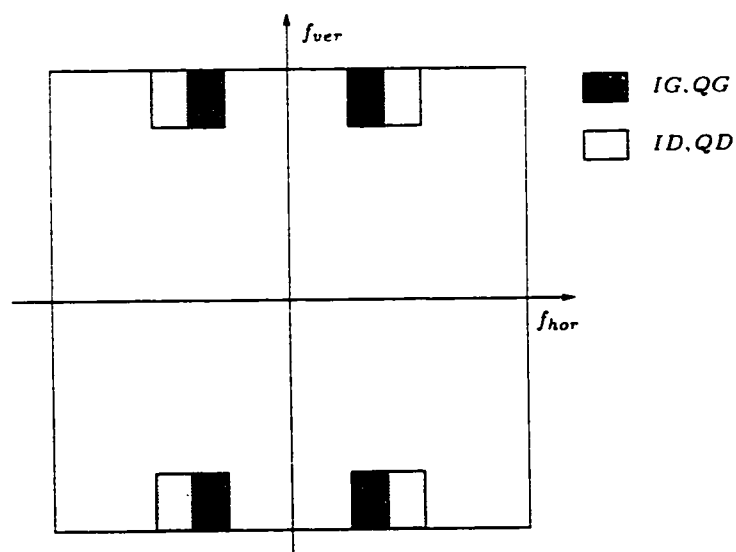


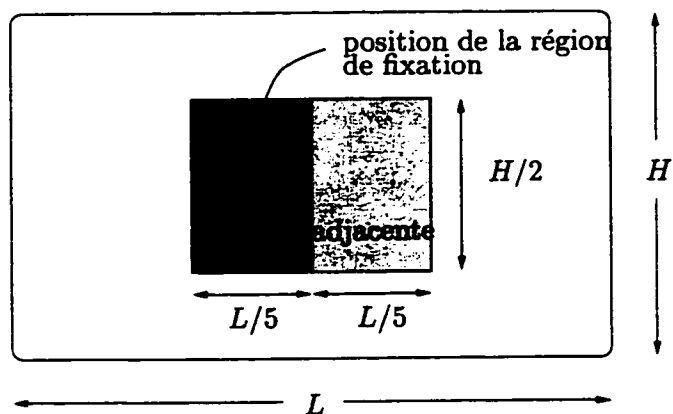
Figure 5.26 : Occupation spectrale des composantes de chrominance avec la méthode *diffusion*. La largeur de bande des composantes $IPBG$ et $IPBD$ est limitée à $f_{sp}/7 = 0,5$ MHz.

struction se trouvent à l'annexe D.

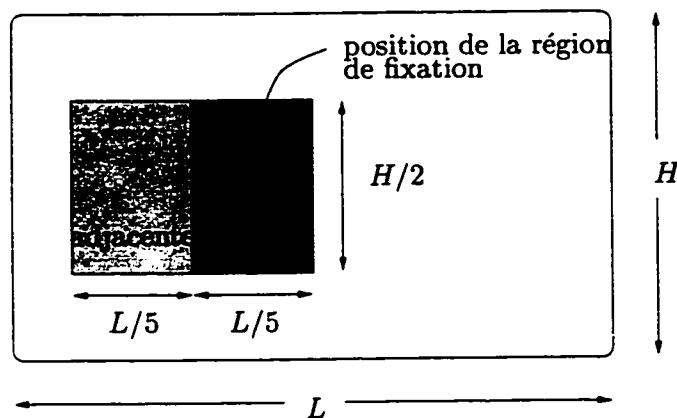
5.4 Méthode intégrée : luminance limitée à $2f_{sp} = 7,16$ MHz ou $f_l = 4,2$ MHz

Il serait certainement utile de posséder une représentation générique d'images qui est à la fois appropriée pour les contraintes de largeur de bande des méthodes *studio* et *diffusion*. Lorsque la bande passante d'un système est assez large pour accommoder un signal de format *studio*, le plus d'information stéréoscopique possible devrait être convoyée par le signal vidéo composite, alors que lorsque la largeur de bande ne peut qu'accueillir un signal de format *diffusion*, le même signal vidéo filtré à $f_l = 4.2$ MHz devrait contenir l'information stéréoscopique essentielle. La méthode de compression selon l'approche *intégrée* suit cette idée en étendant la méthode *diffusion* de façon à y incorporer un traitement de la région de fixation qui tient compte des résultats de l'étude sur les mécanismes de fixation du chapitre 4.

La région de fixation de chaque canal est divisée en deux sous-régions : une sous-région centrale et une sous-région adjacente, tel qu'illustré à la figure 5.27. La taille horizontale de chaque sous-région est égale à un cinquième de la largeur d'un champ et la position de la sous-région centrale correspond à celle de la région de fixation. Pour le point de vue gauche, la sous-région adjacente se trouve à droite de la sous-région centrale, alors que pour le point de vue droit, elle se trouve à gauche de celle-ci. De



(a)



(b)

Figure 5.27 : Division de la région de fixation en 2 sous-régions : (a) point de vue gauche. et (b) point de vue droit.

cette façon, une plus grande partie de la région de fixation se trouve du côté opposé de l'œil qui la regarde pour chaque canal, en conformité avec les résultats de l'étude du chapitre 4.

La partition de l'espace spectral selon la méthode *intégrée* est illustrée à la figure 5.28. L'occupation spectrale jusqu'à f_l est la même que pour la méthode *diffusion*

avec trous de Fukinuki, avec les régions de fixation de la méthode *diffusion* correspondant aux sous-régions centrales de la figure 5.27. L'espace spectral entre f_l et $2f_{sp}$ est utilisé pour insérer de l'information de luminance de hautes-fréquences correspondant aux sous-régions adjacentes de la figure 5.27, pour étendre l'espace spectral de chrominance jusqu'à $3f_{sp}/2$ et pour améliorer la qualité de la composante de luminance passe-bas droite. La sous-région adjacente de chaque canal est traitée d'une façon similaire à celle utilisée pour la sous-région centrale (voir la section 5.3) : la composante passe-haut est divisée en deux sous-composantes représentant les hautes-fréquences verticales supérieures à $f_v/2$ et horizontales inférieures à $f_l/2$, $YPH3G$ et $YPH3D$, et les hautes-fréquences horizontales supérieures à $f_l/2$, $YPH4G$ et $YPH4D$. Ces sous-composantes sont agrandies horizontalement par un facteur de 5 et verticalement par un facteur de 2, puis filtrées passe-bas afin d'éliminer les répétitions spectrales. Les composantes de luminance ainsi traitées, appelées $YRF3G$, $YRF3D$, $YRF4G$ et $YRF4D$, sont modulées pour occuper l'espace spectral tel qu'illustré à la figure 5.28. La composante de luminance passe-bas droite $YPBD$ est maintenant modulée avec double bande latérale de façon à améliorer la qualité de reconstruction lorsque la bande passante est assez large pour accommoder un signal de format *studio*. L'amplitude des composantes de chrominance est modulée en quadrature de façon à exploiter les trous de Fukinuki (voir la figure 5.9). Lorsque la largeur de bande est limitée à f_l , l'information de hautes-fréquences correspondant aux sous-régions adjacentes n'est pas transmise, la composante de luminance passe-bas droite est modulée

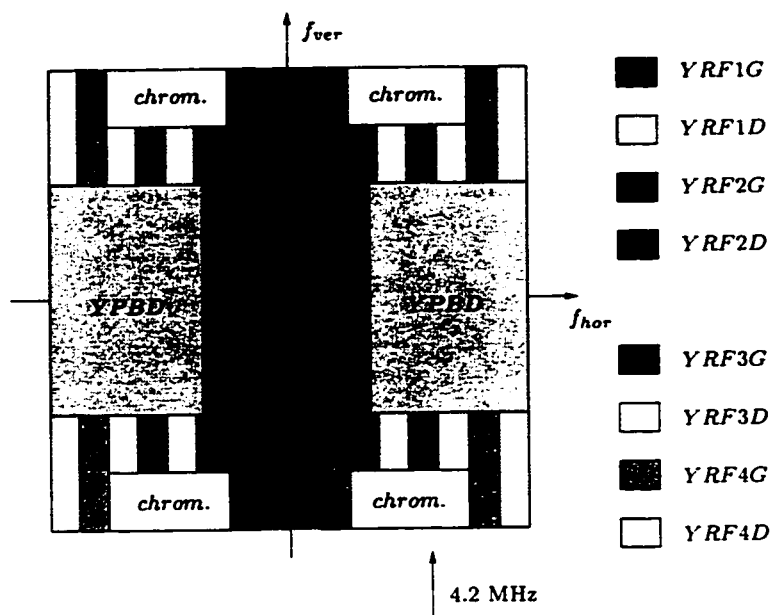


Figure 5.28 : Division de l'espace spectral avec la méthode *intégrée*. $\{YPBG, YPBD\}$, $\{YRF1G, YRF1D, YRF2G, YRF2D, YRF3G, YRF3D, YRF4G, YRF4D\}$ et *chrom.* représentent respectivement les composantes de luminance passe-bas gauche et droite, les sous-composantes passe-haut gauche et droite associées à la région de fixation et les composantes de chrominance.

avec bande latérale résiduelle (VSB) et les composantes de chrominance sont modulées en quadrature avec bande latérale résiduelle (VSB) (voir la figure 5.23), comme pour la méthode *diffusion* avec trous de Fukinuki. La formation du signal vidéo composite selon la méthode *intégrée* est illustré à l'aide du schéma bloc de la figure 5.29 et le traitement détaillé des composantes de luminance se trouve aux figures 5.30, 5.31, 5.32 et 5.33.

La reconstruction des images au récepteur se fait selon les mêmes principes que ceux des méthodes *studio* et *diffusion*. Les schémas bloc illustrant de façon détaillée cette reconstruction se trouvent à l'annexe D.

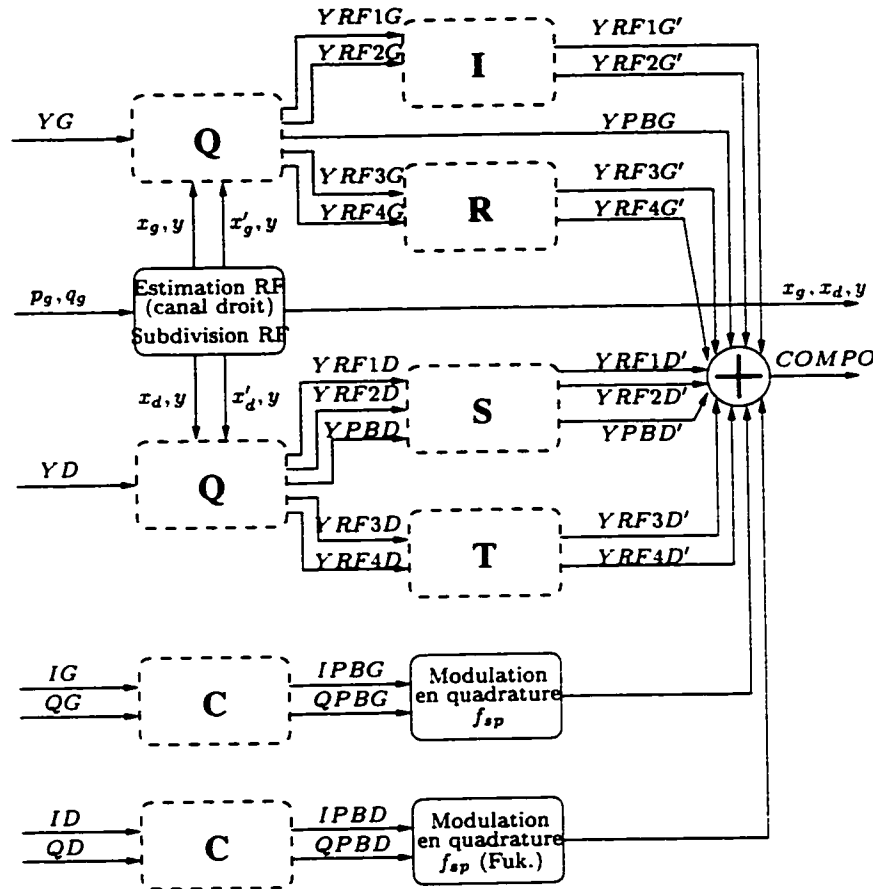


Figure 5.29 : Formation du signal vidéo composite avec la méthode *intégrée*. $\{YG, YD\}$ et $\{IG, QG, ID, QD\}$ correspondent aux composantes originales de luminance et de chrominance des canaux gauche et droit et *COMPO* au signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation dans le canal gauche, alors que x_g et x_d représentent la position de la sous-région de fixation centrale dans les canaux gauche et droit et x'_g et x'_d celle de la sous-région de fixation adjacente dans ces mêmes canaux, avec $y = q_g$ pour les deux canaux. Le bloc **I** se trouve à la figure 5.18, le bloc **Q** à la figure 5.30, le bloc **R** à la figure 5.31, le bloc **S** à la figure 5.32, le bloc **T** à la figure 5.33 et le bloc **C** à la figure 5.5

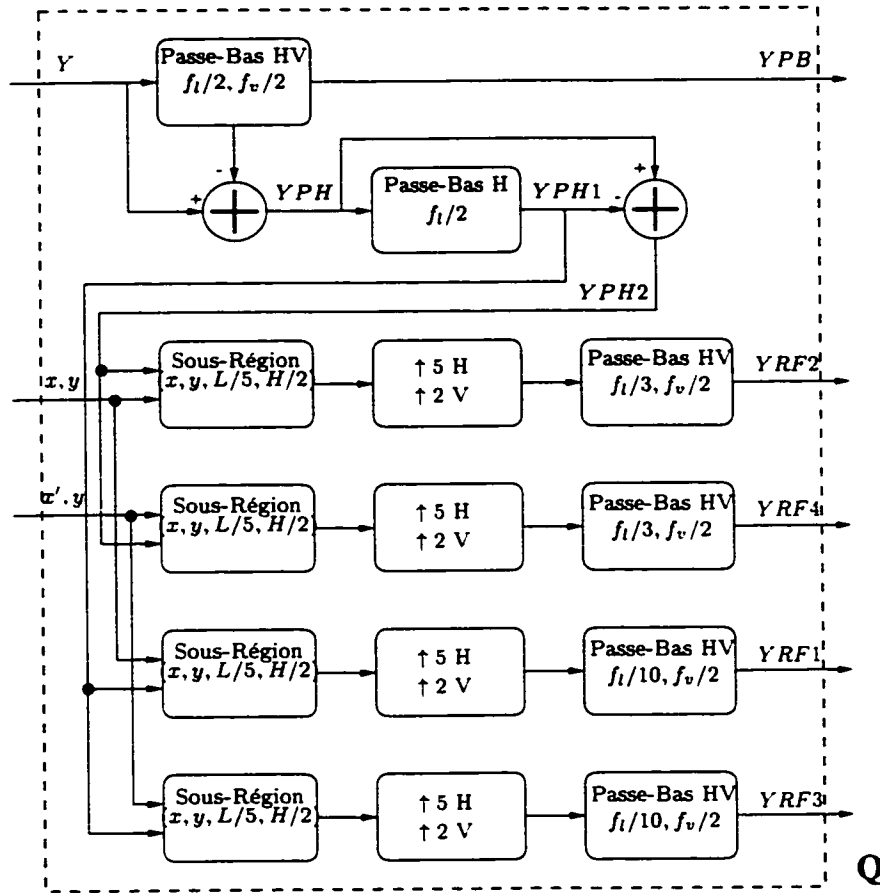


Figure 5.30 : Décomposition de l'information de luminance en composantes passe-bas et passe-haut avec la méthode *intégrée*. Y , YPH , YPB , $\{YRF1, YRF2, YRF3$ et $YRF4\}$ représentent respectivement les composantes de luminance originale, passe-haut, passe-bas et les sous-composantes de luminance passe-haut associée à la région de fixation. x , x' représentent la position horizontale des sous-régions de fixation centrale et adjacente et y la position verticale de ces sous-régions.

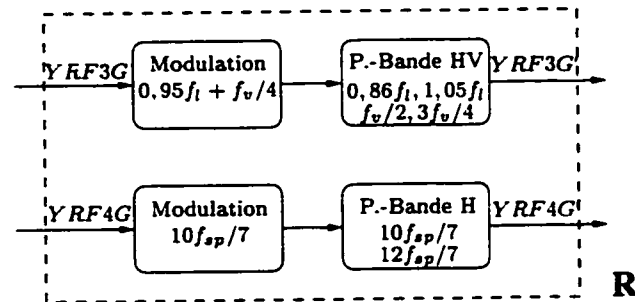


Figure 5.31 : Traitement des composantes passe-haut de la sous-région adjacente du canal gauche avec la méthode *intégrée*. $\{YRF3G, YRF4G\}$, $\{YRF3G', YRF4G'\}$ représentent les deux sous-composantes de luminance passe-haut associées à la sous-région de fixation adjacente et ces sous-composantes modulées.

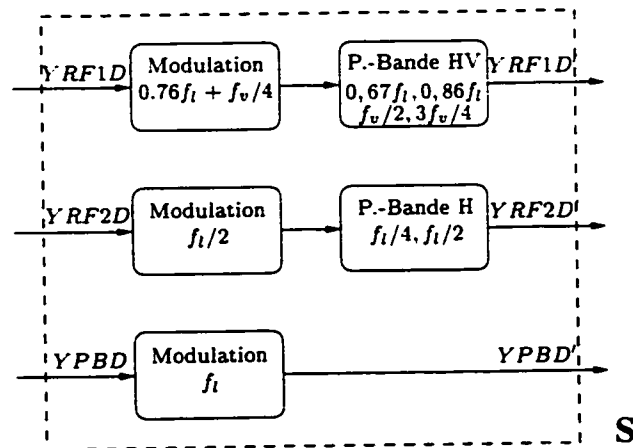


Figure 5.32 : Traitement de la composante passe-bas et des composantes passe-haut de la sous-région centrale du canal droit avec la méthode *intégrée*. $YPBD$, $\{YRF1D, YRF2D\}$, $YPBD'$, $\{YRF1D'$ et $YRF2D'\}$ représentent la composante de luminance passe-bas, les deux sous-composantes de luminance passe-haut associées à la sous-région de fixation centrale et ces trois composantes modulées.

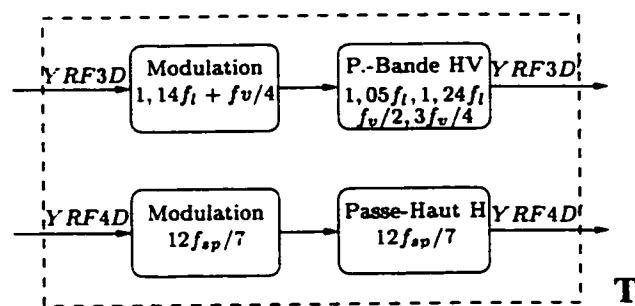


Figure 5.33 : Traitement des composantes passe-haut de la sous-région adjacente du canal droit avec la méthode *intégrée*. $\{YRF3D, YRF4D\}$, $\{YRF3D'$ et $YRF4D'\}$ représentent les deux sous-composantes de luminance passe-haut associées à la sous-région de fixation adjacente et ces mêmes sous-composantes modulées.

5.5 Résumé des 3 méthodes

Les figures 5.34. 5.35 et 5.36 résument les principales étapes de traitement des 3 méthodes de codages des sections précédentes et les tableaux 5.1 et 5.2 présentent les caractéristiques les plus importantes de ces méthodes. Dans ces tableaux, la largeur de bande et l'occupation spectrale sont données en MHz, L et H correspondent respectivement à la largeur et la hauteur d'un champ, le nombre de (sous-) composantes est donné pour chaque canal stéréoscopique, et BLR et Fuk. correspondent respectivement à une modulation avec bande latérale résiduelle et avec trous de Fukinuki. Les données des tableaux 5.1 and 5.2 montrent clairement que la méthode *intégrée* représente un compromis entre les méthodes *studio* et *diffusion*. En effet, pour une largeur de bande de $2f_{sp} = 7.16$ MHz, la largeur de bande avec la méthode *intégrée* est de $2L/5$ et la largeur de bande des composantes de luminance passe-bas est de $f_l/2$, alors que ces paramètres sont égaux à $L/2$ et f_{sp} avec la méthode *studio*. Cette

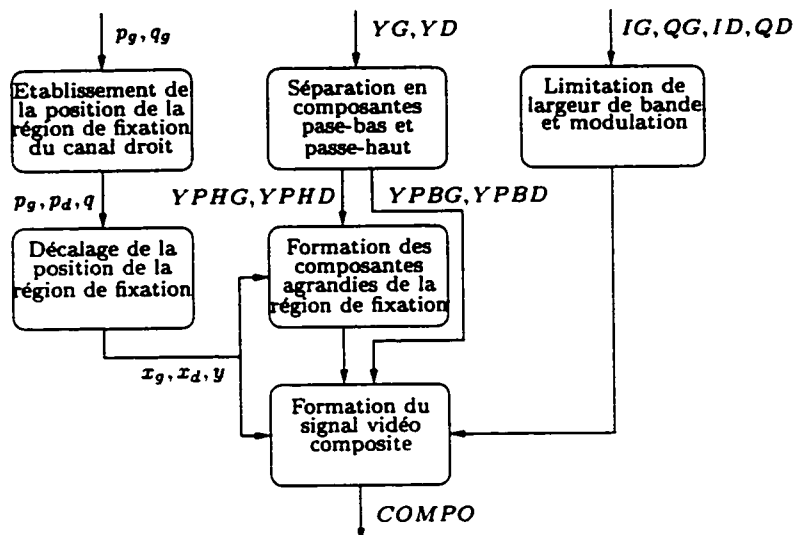


Figure 5.34 : Principales étapes de traitement avec la méthode *studio*. $\{Y_G, Y_D\}$, $\{I_G, Q_G, I_D, Q_D\}$ et $COMPO$ représentent respectivement les composantes de luminance et de chrominance et le signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation pour le canal gauche, alors que x_g et x_d représentent la position décalée de la région de fixation dans les canaux gauche et droit, avec $y = q_g$ pour les deux canaux.

réduction de largeur de bande et de taille de la région de fixation est toutefois compensée par une flexibilité accrue qui rend un signal vidéo de format *intégré* adéquat autant pour des applications où la largeur de bande est de $2f_{sp}$ que pour celles où elle est limitée à f_l .

La représentation par composantes utilisées par les méthodes de ce chapitre représente une façon efficace de compacter l'information à transmettre puisque l'information essentielle de haute-fréquences (correspondant à la région de fixation) occupe un espace spectral réduit. La combinaison des composantes par manipulations spectrales afin de former un signal vidéo composite approprié pour une transmission

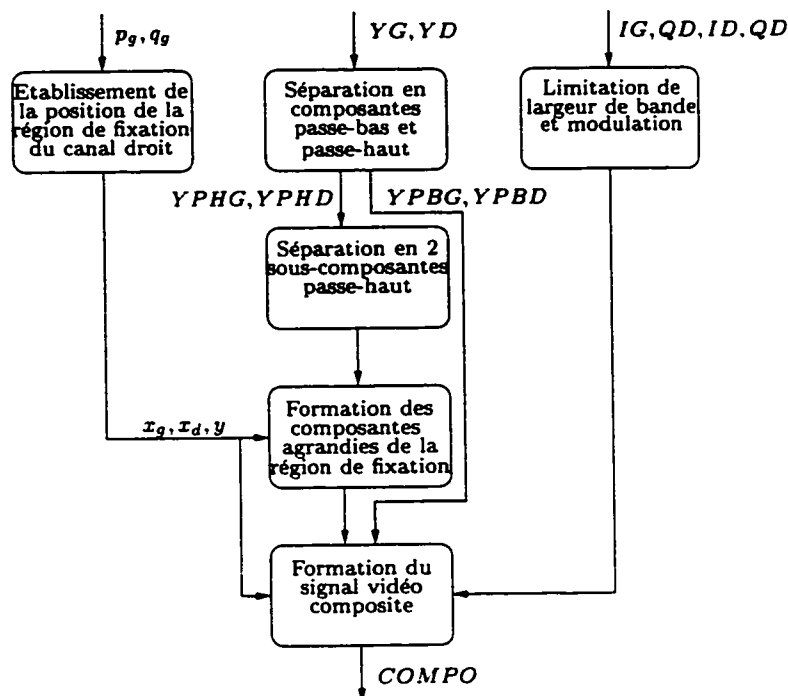


Figure 5.35 : Principales étapes de traitement avec la méthode *diffusion*. $\{YG, YD\}$, $\{IG, QG, ID, QD\}$ et $COMPO$ représentent respectivement les composantes de luminance et de chrominance et le signal vidéo composite. p_g et q_g représentent la position estimée de la région de fixation pour le canal gauche : $x_g = p_g$, x_d représente la position de la région de fixation dans le canal droit et $y = q_g$ pour les deux canaux.

Tableau 5.1 : Caractéristiques du traitement de luminance avec les différentes méthodes de codage.

	studio	diffusion	intégrée
Largeur région fixation	$L/2$	$L/5$	$L/5$ ou $2L/5$
Hauteur région fixation	$H/2$	$H/2$	$H/2$
Traitement région fixation	décalage	aucun	2 sous-régions
Largeur bande comp. p.-bas	3.58	2.1	2.1
Sous-comp. p.-haut	non	oui	oui
(Sous-) comp. p.-haut	1	2	2 or 4
Occupation spectrale	7.16	4.2	4.2 or 7.16

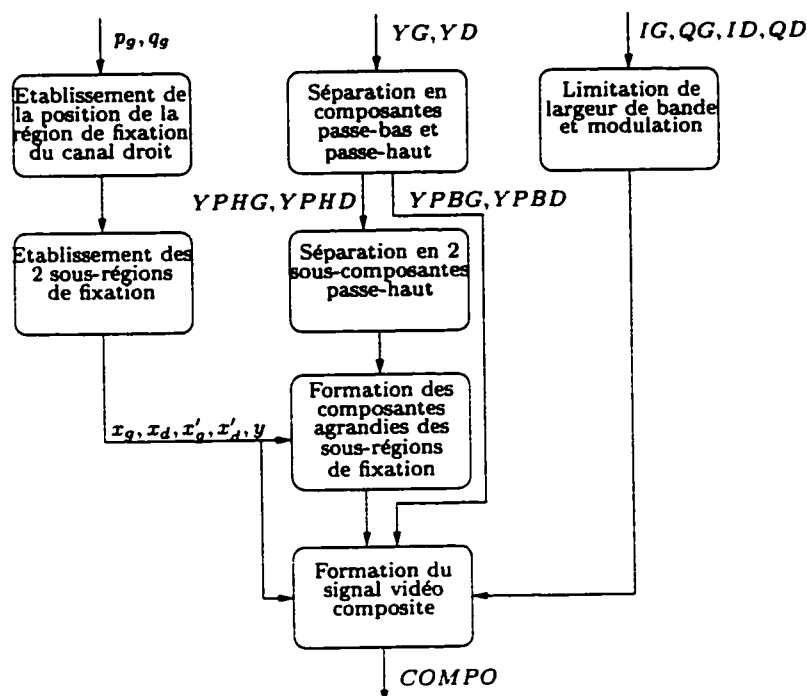
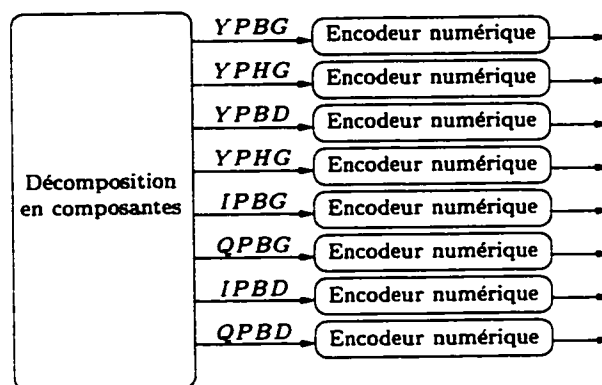


Figure 5.36 : Principales étapes de traitement avec la méthode *intégrée*. $\{YG, YD\}$, $\{IG, QG, ID, QD\}$ et $COMPO$ correspondent respectivement aux composantes de luminance et de chrominance et au signal vidéo composite. pg et qg représentent la position estimée de la région de fixation dans le canal gauche, alors que xg et xd représentent la position de la sous-région de fixation centrale dans les canaux gauche et droit et xg' et xd' celle de la sous-région de fixation adjacente dans ces canaux, avec $x_g = p_g$ et $y = q_g$ pour les deux canaux.

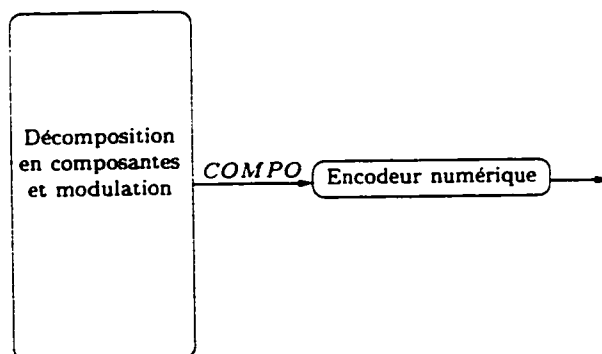
Tableau 5.2 : Caractéristiques du traitement de chrominance avec les différentes méthodes de codage.

	Largeur de bande I	Largeur de bande Q
studio	1,5 (BLR)	0,5
studio (Fuk.)	1,5	0,5
diffusion	0,5	0,5
diffusion (Fuk.)	1,5 (BLR)	0,5
intégrée (7.16)	1,5	0,5
intégrée (4.2)	1,5 (BLR)	0,5

d'information sous forme analogique représente une façon possible d'exploiter cette représentation. Il devrait néanmoins être clair que la représentation par composante originale qui est proposée est aussi adéquate pour une transmission numérique de l'information. A cette fin, la représentation par composante serait utilisée comme une étape de pré-traitement avant l'utilisation d'un (ou plusieurs) encodeur(s) numérique(s) standard, tel qu'un encodeur MPEG ou JPEG. L'information fournie à l'entrée de cet (ces) encodeur(s) peut prendre la forme de composantes individuelles ou d'un signal composite. La figure 5.37 illustre deux façons possibles d'exploiter la représentation par composantes proposée pour une transmission numérique de l'information. Puisque l'information de haute-fréquences a été manipulée de façon à être représentée à l'intérieur d'un espace spectral limité, il est prévisible que toutes les composantes de la figure 5.37.a seront codées à l'aide d'une représentation numérique compacte. A la figure 5.37.b, la représentation de l'information avec un signal vidéo composite implique déjà une compression 4:1 puisque l'information de luminance et de chrominance des deux canaux est représentée avec le même nombre de bits qu'une composante de luminance originale. Le taux de compression sera par la suite augmenté en fonction du traitement accompli par l'encodeur numérique se trouvant à cette figure.



(a)



(b)

Figure 5.37 : Transmission numérique de l'information : (a) en codant les composantes individuellement, et (b) en codant le signal vidéo composite. $\{YPBG, YPBD, YPHG, YPHD\}$, $\{IG, QG, ID, QD\}$ et *COMPO* représentent respectivement les composantes de luminance passe-bas et passe-haut, les composantes de chrominance passe-bas et le signal vidéo composite. Les composantes de luminance passe-haut ont été limitées à une région de fixation et agrandies (zoom) de façon à réduire leur occupation spectrale.

Chapitre 6

Les résultats expérimentaux

Cinq séquences stéréoscopiques ont été codées avec les méthodes décrites au chapitre 5 et la qualité des résultats obtenus a été évaluée de façon subjective par quelques sujets et de façon objective selon une mesure de rapport signal à bruit entre les images originales et reconstruites. Pour l'évaluation subjective, trois sujets ont noté la qualité des images codées selon une échelle de visibilité de dégradations. L'évaluation objective a porté sur les trois principales étapes de traitement nécessaires à la transmission des images, soit la représentation originale des images en composantes à largeur de bande limitée, la modulation et la démodulation des composantes transmises et la reconstruction des images en recombinaison de ces composantes.

Les séquences qui ont été testées proviennent du Centre Commun d'Etudes de Télédiffusion et Télécommunications (CCETT) en France. Elles ont donc été acquises selon un format CCIR 601 européen, c'est-à-dire 50 champs par seconde avec 288 lignes

actives de 720 pixels, échantillonnés à 13.5 MHz et représentés avec les composantes Y, U et V . Avant d'être codées selon les méthodes du chapitre 5, ces séquences ont tout d'abord été converties au format NTSC en transformant les composantes Y, U et V en composantes Y, I et Q .

Le nombre de lignes actives de chaque champ a été limité à 240 en éliminant les 48 lignes supérieures des champs originaux et une conversion de fréquence d'échantillonnage a été accomplie afin d'obtenir des séquences échantillonnées à $4f_{sp} = 14.32$ MHz (voir Oppenheim et Schafer (1989) pour la conversion de fréquence d'échantillonnage). Après ces opérations, la taille des champs est maintenant de 240 lignes de 760 pixels. Idéalement, une conversion de fréquence temporelle afin de passer de 50 Hz à 60 Hz devrait aussi être accomplie. Toutefois, en raison de sa complexité et du faible impact de cette conversion sur la performance de méthodes de codage du chapitre 5, cette étape de traitement n'a pas été accomplie. Il est assez facile de s'imaginer que, à la place de coder les scènes qui ont été captées à l'origine, des scènes dont le mouvement des objets est légèrement plus rapide de façon à ce qu'un déplacement d'une durée de 1,2 secondes se fasse en 1 seconde sont plutôt utilisées. Les images ainsi converties au format NTSC sont maintenant considérées comme les données originales qui seront traitées par les différentes méthodes de codage du chapitre 5. En conséquence il ne sera plus nécessaire de faire référence aux images en format européen.

La seconde image du canal gauche des cinq séquences qui ont été codées est illustrée à la figure 6.1. Ces séquences contiennent différents types de mouvement et

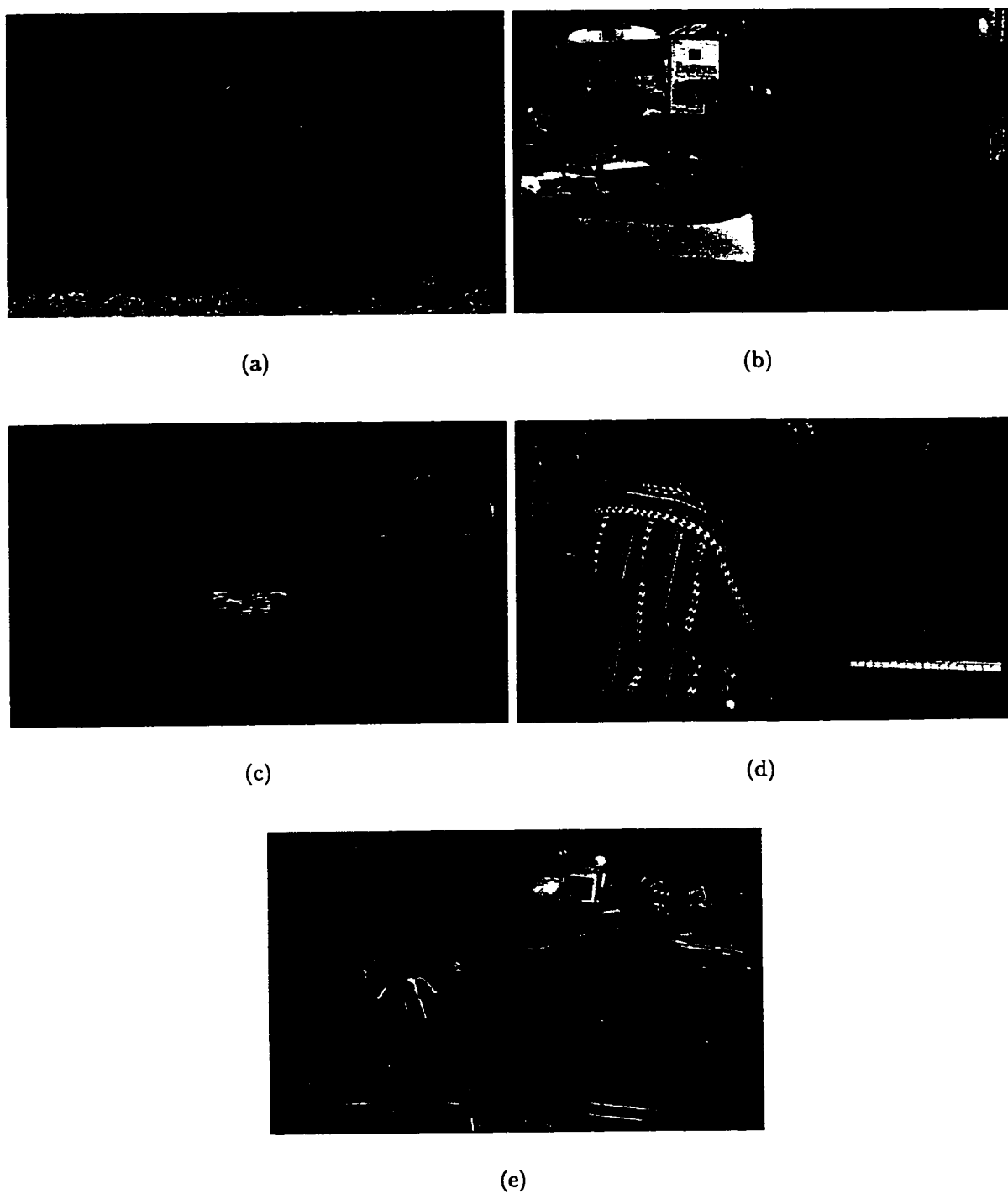


Figure 6.1 : Composantes de luminance originales du canal de gauche de la deuxième image des séquences testées : (a) *aqua*, (b) *manège*, (c) *train*, (d) *piano* et (e) *tunnel*.

plusieurs variations de profondeur. La séquence *aqua* représente un aquarium artificiel avec des pierres, des plantes et des poissons en caoutchouc qui sont suspendus devant une affiche représentant une vue sous-marine. Les poissons ne bougent presque pas et la caméra exécute un lent mouvement latéral. Pour cette séquence, la région de fixation a été choisie sur la nageoire du gros poisson se trouvant en bas à gauche de l'image en raison de la variété des couleurs se trouvant dans cette zone. Puisque la scène contient peu de mouvements, la position de la région de fixation ne change dans l'image que pour compenser le mouvement de la caméra. La séquence *manège* contient un carrousel en région urbaine. Le fond de la séquence représente donc une scène avec des édifices, une route, des piétons et un autobus. A l'avant-plan, la rotation du carrousel fait apparaître et disparaître les différents jouets qu'il contient. Pour cette séquence, la région de fixation a été choisie comme étant le voilier (un des jouets du carrousel) parce qu'il subit un grand déplacement. Même si ce n'est pas l'objet le plus détaillé de la scène, il parcourt en environ une seconde la presque totalité de l'image selon un mouvement de rotation qui l'amène de l'extrême gauche de l'image à la droite de celle-ci. La séquence *train* représente une scène de jouets d'enfant avec deux trains se déplaçant latéralement selon des directions opposées qui se croisent. Les barils sur un des trains ont été choisis comme la région de fixation. Ces barils sont tantôt visibles, tantôt partiellement cachés et même temporairement invisibles en raison des occlusions lors du croisement des trains. La séquence *piano* représente un mur uniforme bleu devant lequel un pianiste vu de dos est en train de

jouer. Il porte une chemise à motifs texturés noirs et blancs qu'il fait onduler en fonction de ses mouvements. La région de fixation a été choisie comme étant la chemise en raison des fins détails de ses motifs. Le mouvement de la chemise et aussi celui des doigts du pianiste peuvent être considérés comme des mouvements non-rigides. La séquence *tunnel* représente une scène de jouets d'enfants avec un train qui entre dans un tunnel en construction en suivant une voie ferrée courbée. Le mouvement d'éloignement du train cause une déformation de la forme apparente des wagons. À côté du tunnel, une équipe de cinéma tourne un film. La caméra est fixe et le seul mouvement de la scène est celui du train. Le personnage se trouvant devant le cheval a été choisi comme région de fixation car il semble que le focus des caméras soit fait sur l'endroit où il se trouve. Puisque la caméra est fixe, la position de la région de fixation ne change pas durant la séquence.

6.1 Illustration des étapes de traitement

Afin d'illustrer les résultats obtenus, les composantes d'une des images d'une séquence codée sont présentées à différentes étapes de traitement avec les méthodes *studio*, *diffusion* et *intégrée*. Les composantes présentées sont l'entrelacement des troisième et quatrième champs du canal droit de la séquence *aqua*. Pour cette séquence, la région de fixation se trouve sur le gros poisson situé en bas et à gauche de l'image, plus spécifiquement sur sa nageoire qui apparaît comme une région ovale derrière et sous la tête du poisson (voir la figure 6.2). La figure 6.2 illustre la com-

posante de luminance originale et la figure 6.3 les composantes de luminance passe-bas originale et reconstruite avec la méthode *studio*. En raison de la résolution limitée de l'imprimante, les différences entre ces trois images peuvent être difficiles à percevoir sur papier, bien qu'elles soient clairement visibles lorsque les images sont montrées sur un écran d'ordinateur. La figure 6.4 montre les composantes de luminance passe-haut de la région de fixation originale et reconstruite pour la méthode *studio*. Pour ces images ainsi que toutes les images subséquentes représentant de l'information passe-haut ou de chrominance, une valeur égale à 0 est représenté par un niveau de gris intermédiaire et les valeurs négatives et positives par des niveaux de gris respectivement plus foncés et plus pâles que le niveau intermédiaire. Afin de faciliter la visualisation de ces images sur papier, une technique d'amélioration de contraste a été utilisée. Cette technique consiste à *étirer* l'intervalle des niveaux de gris intermédiaires correspondant aux petites différences de façon à augmenter leur visibilité (Jain 1989). Il faut observer que la région de fixation n'est pas centrée par rapport au point de fixation, mais bien décalée vers la gauche par rapport à ce point afin de tenir compte des résultats de l'étude psychophysique de chapitre 4. Une reconstruction parfaite de l'information passe-haut n'est certainement pas espérée avec les méthodes proposées, mais il est clair que l'information passe-haut de la composante reconstruite est similaire à celle transmise, tel que désiré. La composante passe-haut qui est utilisée dans la formation de signal vidéo n'est pas celle de la figure 6.4, mais cette composante agrandie par un facteur de 2 dans les directions horizontale et verticale. Cette com-

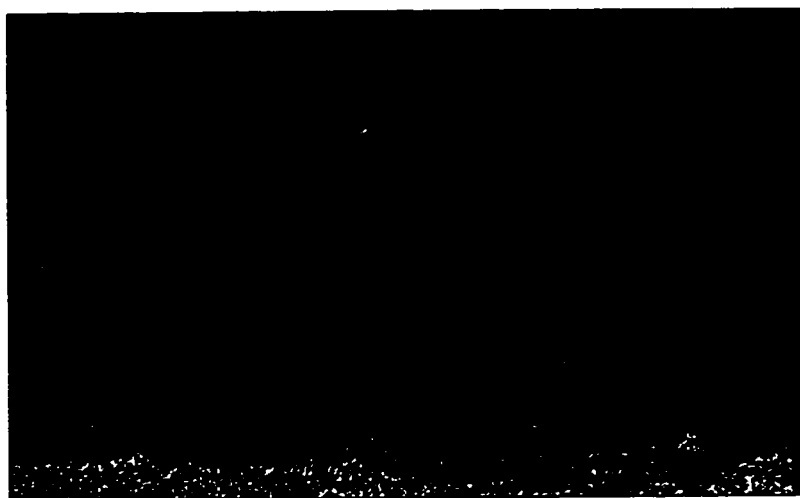
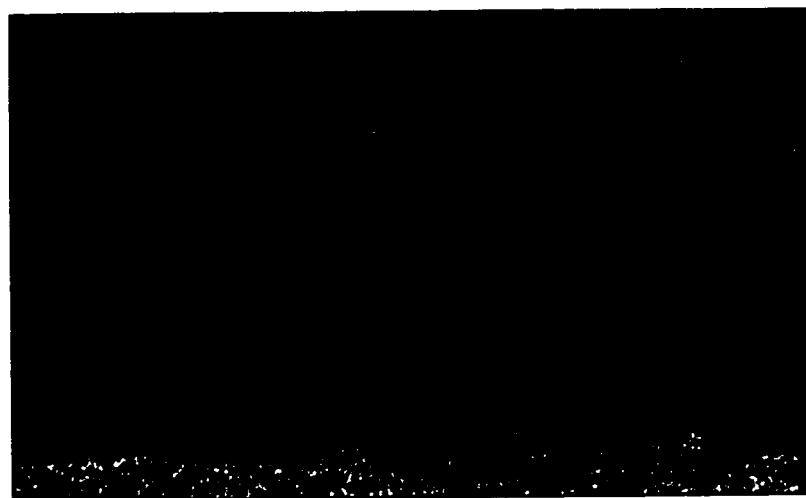


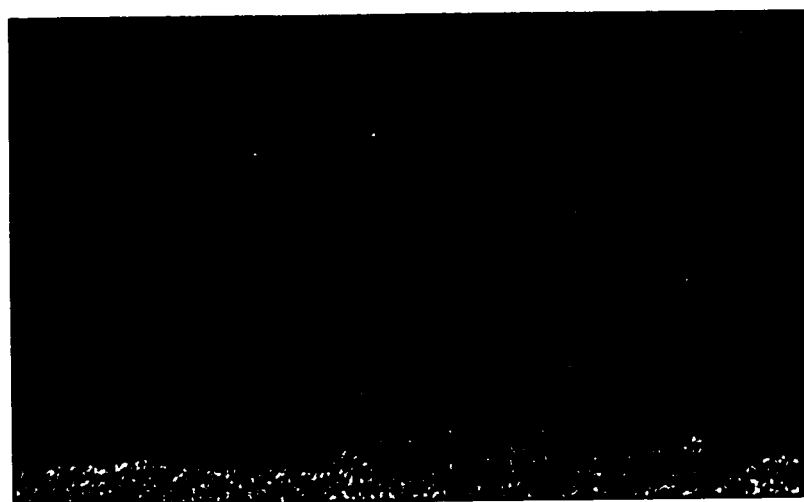
Figure 6.2 : Composante de luminance originale de la deuxième image de la séquence *aqua*.

posante est illustrée à la figure 6.5 avant la formation du signal composite et suite à un filtrage pour la séparation des différentes composantes du signal vidéo. Il est assez facile de reconnaître la forme du poisson dans ces images malgré la dilatation de celles-ci. La figure 6.6 illustre la composante de luminance reconstruite qui est obtenue en ajoutant la composante de la figure 6.4.b à celle de la figure 6.3.b. En raison de la résolution limitée de l'imprimante, il peut être difficile de voir la différence entre la composante reconstruite de la figure 6.6 et la composante originale de la figure 6.2. Afin de faciliter cette visualisation, la figure 6.7 présente la différence entre les composantes de luminance reconstruite et originale. Dans cette figure, l'amplitude des valeurs de différences est plus petite dans la région de fixation, tel que désiré.

Les figures 6.8 et 6.9 illustrent les composantes de chrominance I et Q , limitées respectivement à $3f_{sp}/7 = 1,5$ MHz et $f_{sp}/7 = 0,5$ MHz, avant et après leur transmission.



(a)



(b)

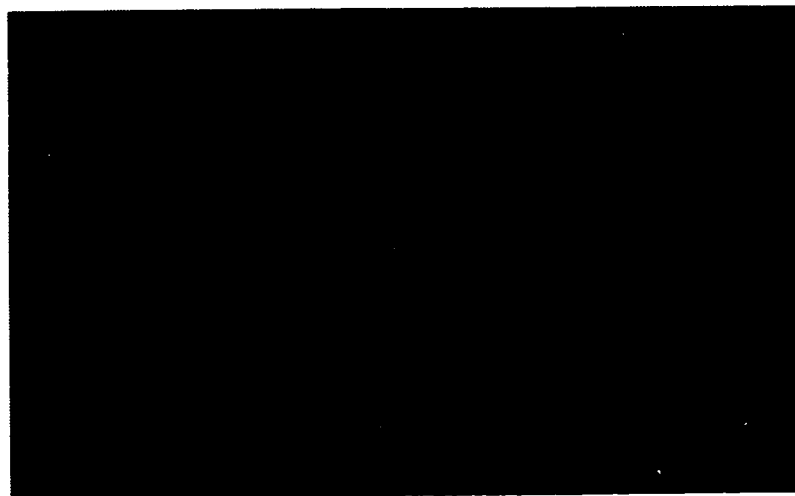
Figure 6.3 : Composantes de luminance passe-bas de la deuxième image de la séquence *aqua* avec la méthode *studio* : (a) composante originale et (b) composante reconstruite.



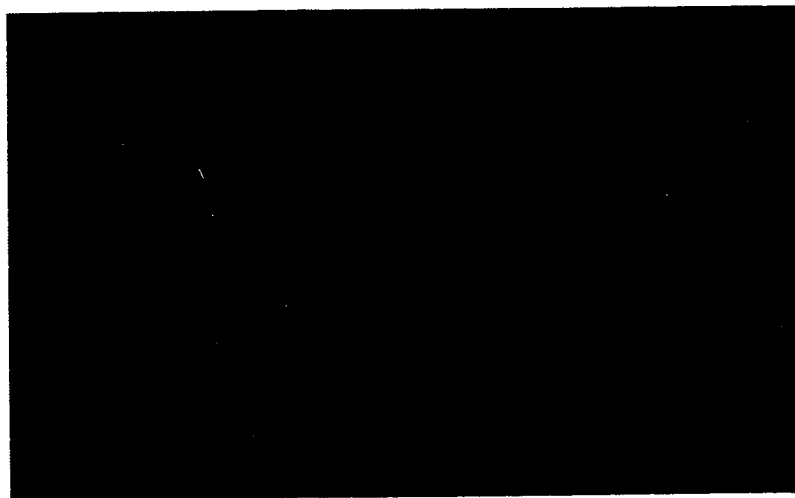
Figure 6.4 : Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence *aqua* avec la méthode *studio* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

La figure 6.10 présente le signal vidéo composite obtenu avec la méthode *studio*. Le contenu de la scène est clairement identifiable à partir de ce signal, ce qui laisse présager que, pour une transmission numérique de l'information, les algorithmes de compression utilisés par certains encodeurs numériques devraient permettre une compression additionnelle du signal (voir la discussion à la fin du chapitre 5).

La figure 6.11 présente les composantes de luminance passe-bas originale et reconstruite selon la méthode *diffusion*. Même s'il peut être difficile de le voir en raison de la résolution limitée de l'imprimante, l'information se trouvant dans cette figure est moins détaillée que celle de la figure 6.3 puisqu'elle est limitée horizontalement à $f_l/2 = 2.1$ MHz plutôt qu'à $f_{sp} = 3.58$ MHz. La figure 6.12 illustre la composante passe-haut de la région de fixation. Cette figure combine les deux sous-composantes d'information passe-haut qui sont utilisées avec la méthode *diffusion*, soit les hautes-fréquences verticales dont la fréquence horizontale est plus petite que $f_l/2$ et les



(a)



(b)

Figure 6.5 : Composantes de luminance passe-haut agrandies de la région de fixation de la deuxième image de la séquence *aqua* avec la méthode *studio* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

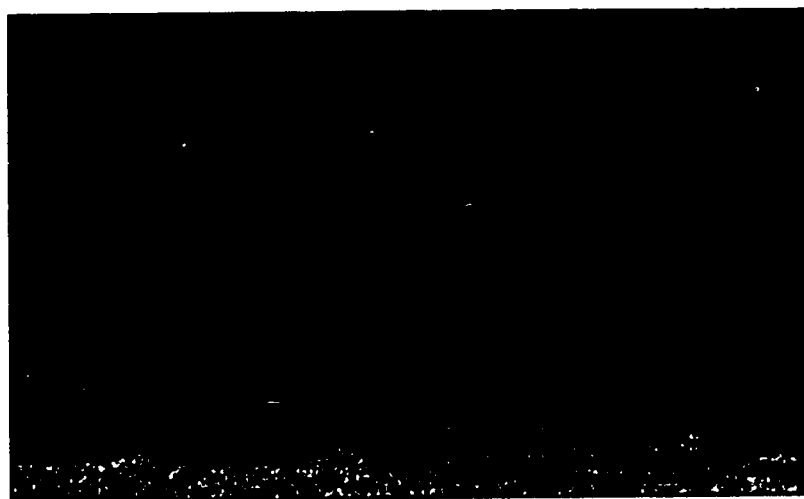


Figure 6.6 : Composante de luminance reconstruite de la deuxième image de la séquence *aqua* avec la méthode *studio*.

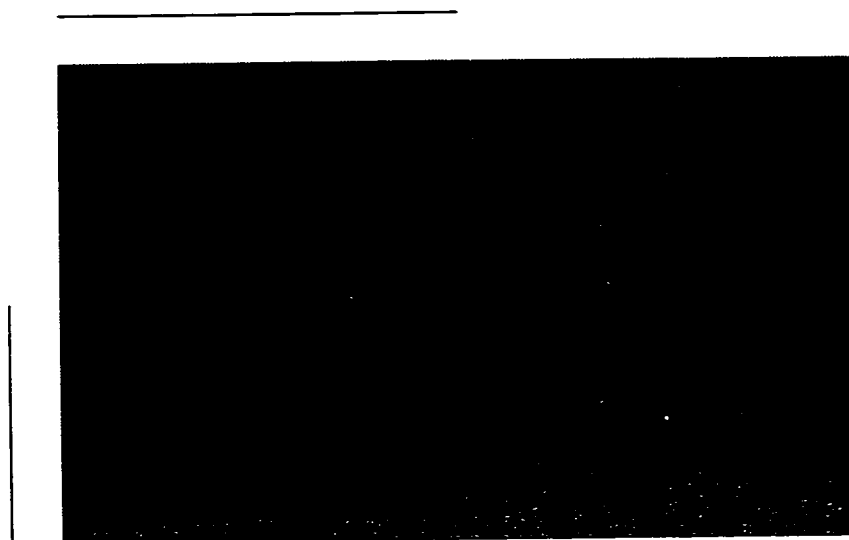


Figure 6.7 : Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence *aqua* avec la méthode *studio*. Les lignes horizontale et verticale en bordure de l'image indiquent la région de fixation. Des techniques d'amélioration de contraste ont été utilisées afin de faciliter la visualisation sur papier.

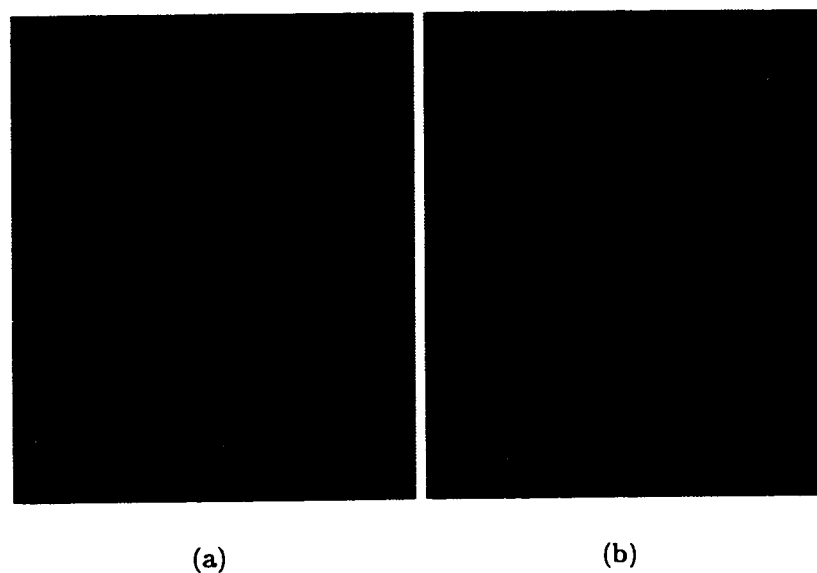


Figure 6.8 : Composantes de chrominance I de la deuxième image de la séquence *aqua* avec la méthode *studio* : (a) composante originale et (b) composante reconstruite.

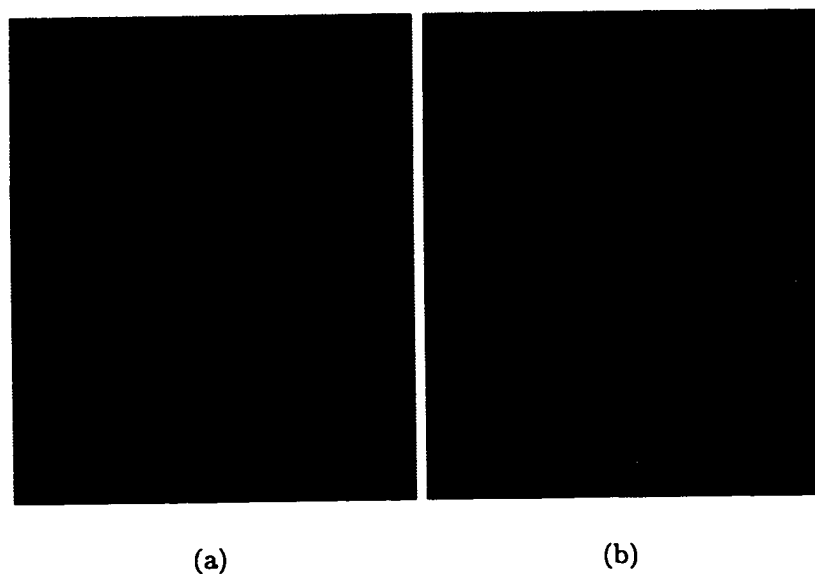


Figure 6.9 : Composantes de chrominance Q de la deuxième image de la séquence *aqua* avec la méthode *studio* : (a) composante originale et (b) composante reconstruite.

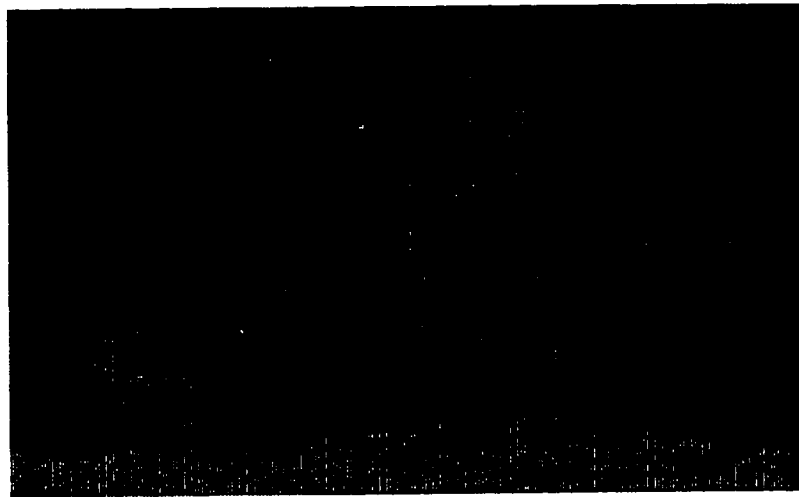


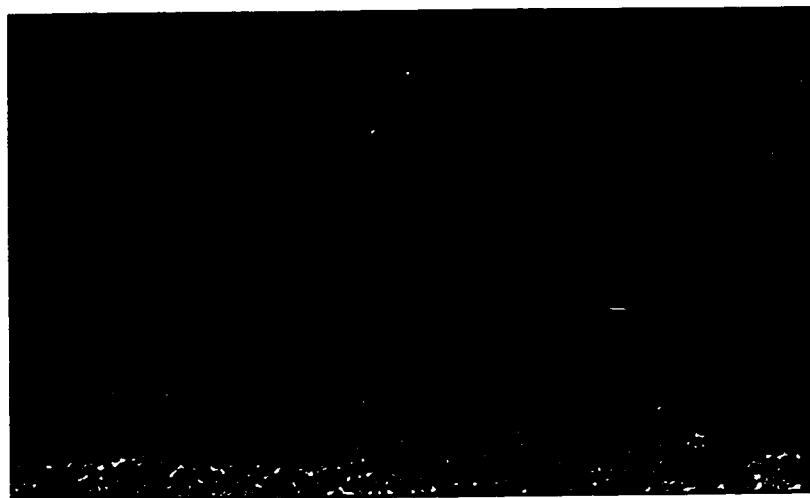
Figure 6.10 : Signal vidéo composite obtenu avec la méthode *studio*.

hautes-fréquences horizontales plus élevées que $f_1/2$. Avec la méthode *diffusion* la taille horizontale de la région de fixation est plus petite qu'avec la méthode *studio* puisqu'elle correspond à un cinquième de la largeur de l'image plutôt qu'à la moitié de celle-ci. La figure 6.13 illustre les composantes passe-haut agrandies originale et reconstruite de la région de fixation. Dans ces images la région de fixation est dilatée par un facteur de 5 dans la direction horizontale et par un facteur de 2 dans la direction verticale. La figure 6.14 illustre la composante de luminance reconstruite qui est obtenue en ajoutant la composante de la figure 6.12.b à celle de la figure 6.11.b. La figure 6.15 présente la différence entre les composantes de luminance reconstruite et originale. Dans cette figure, il peut être plus difficile de constater que l'amplitude des valeurs de différences est bel et bien plus petite dans la région de fixation en raison de sa taille beaucoup plus petite que celle de la méthode *studio*. Néanmoins,

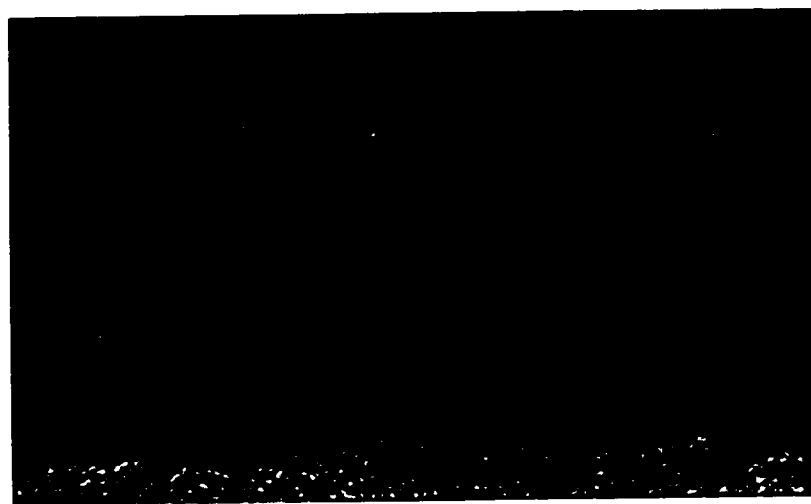
en regardant attentivement dans la région autour de la nageoire du gros poisson se trouvant en bas à gauche de l'image, il est possible de vérifier que l'amplitude des valeurs de différence est effectivement plus petite dans cette région.

Les composantes de chrominance originales qui sont utilisées par la méthode *diffusion* avec trous de Fukinuki sont illustrées aux figures 6.8.a et 6.9.a. Les composantes de chrominance reconstruites selon cette méthodes sont similaires à celles des figures 6.8.b et 6.9.b (compte tenu de la résolution de l'imprimante) et conséquemment elles ne sont pas reproduites une autre fois. Lorsque la méthode *diffusion* sans trous de Fukinuki est utilisée, la composante de chrominance I est limitée à $f_{sp}/7 = 0,5$ MHz. Cette composante est illustrée à la figure 6.16.

Pour la méthode *intégrée* avec une largeur de bande limitée à $f_l = 4.2$ MHz, les composantes de luminance et de chrominance sont les mêmes que celles pour la méthode *diffusion* avec trous de Fukinuki. Pour la même méthode avec une largeur de bande de $2f_{sp} = 7.16$ MHz, les composantes de luminance passe-bas et de chrominance originales sont les mêmes que celles illustrées aux figures 6.11.a, 6.8.a et 6.9.a et ces composantes reconstruites sont similaires à celles des figures 6.11.b, 6.8.b et 6.9.b. Les composantes de luminance passe-haut originale et reconstruite de la région de fixation sont présentées à la figure 6.17. La taille de la région de fixation est plus grande qu'avec la méthode *diffusion* (voir la figure 6.12) et s'étend vers la gauche du point de fixation de façon à tenir compte des résultats de l'étude psychophysique du chapitre 4. La composante de luminance passe-haut reconstruite de la figure 6.17 a



(a)



(b)

Figure 6.11 : Composantes de luminance passe-bas de la deuxième image de la séquence *aqua* avec la méthode *diffusion* : (a) composante originale et (b) composante reconstruite.

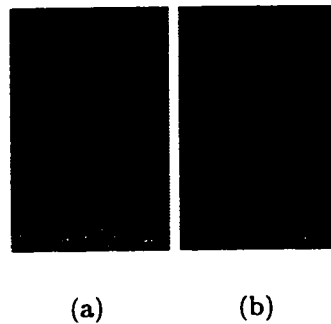
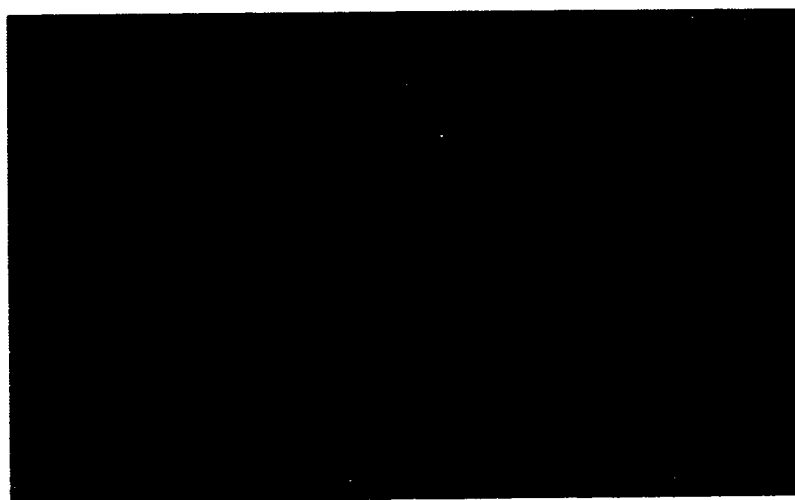


Figure 6.12 : Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence *aqua* avec la méthode *diffusion* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

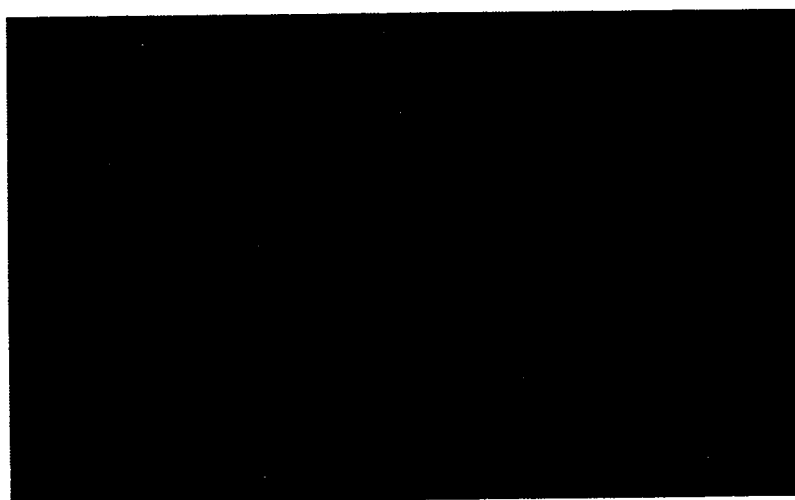
été obtenue à partir de la combinaison des deux sous-régions agrandies des figures 6.18 et 6.19. Finalement, la figure 6.20 illustre la composante de luminance reconstruite qui est obtenue en ajoutant la composante de la figure 6.17.b. à celle de luminance passe-bas reconstruite et la figure 6.21 illustre la différence entre les composantes de luminance reconstruite et originale. Dans cette figure, l'amplitude des différences est plus petite dans la région de fixation, tel que désiré.

6.2 Evaluation subjective

Trois sujets ont évalué la qualité des images codées par les différentes méthodes du chapitre 5 en utilisant l'échelle de visibilité de dégradations du tableau 6.1. Cette échelle est couramment utilisée dans le domaine du traitement d'images (Jain 1989). En plus de noter les séquences selon cette échelle, les sujets devaient aussi consigner



(a)



(b)

Figure 6.13 : Composantes de luminance passe-haut agrandies de la région de fixation de la deuxième image de la séquence *aqua* avec la méthode *diffusion* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

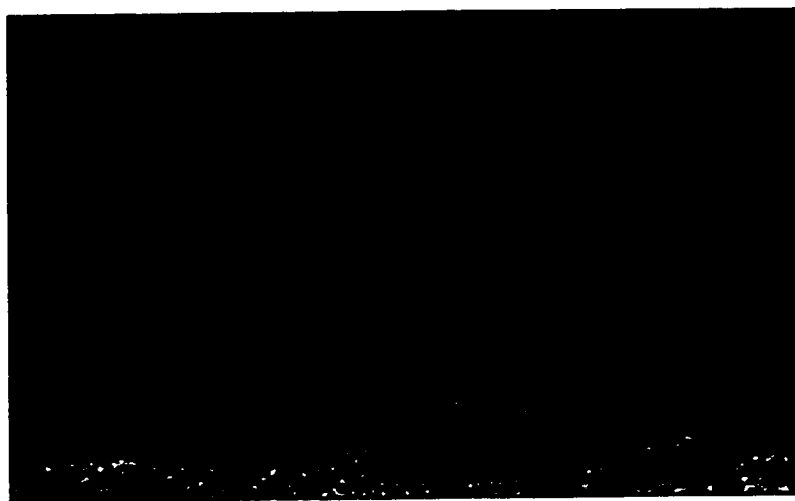


Figure 6.14 : Composante de luminance reconstruite de la deuxième image de la séquence *aqua* avec la méthode *diffusion*.

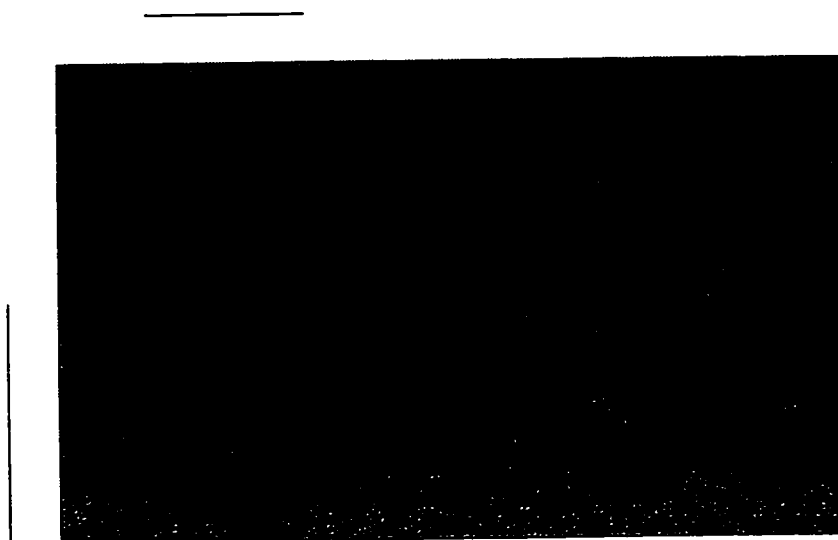


Figure 6.15 : Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence *aqua* avec la méthode *diffusion*. Les lignes horizontale et verticale en bordure de l'image indiquent la région de fixation. Des techniques d'amélioration de contraste ont été utilisées afin de faciliter la visualisation sur papier.

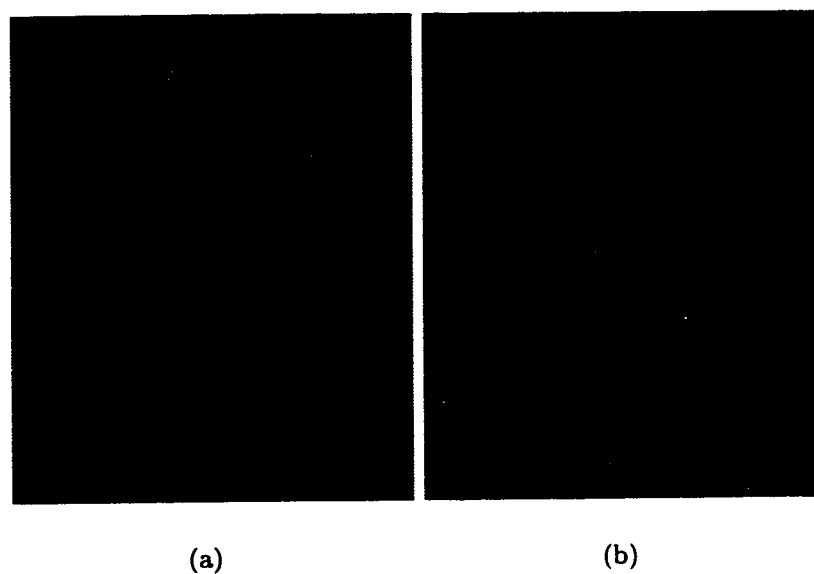


Figure 6.16 : Composantes de chrominance I de la deuxième image de la séquence *aqua* avec la méthode *diffusion* : (a) composante originale et (b) composante reconstruite.

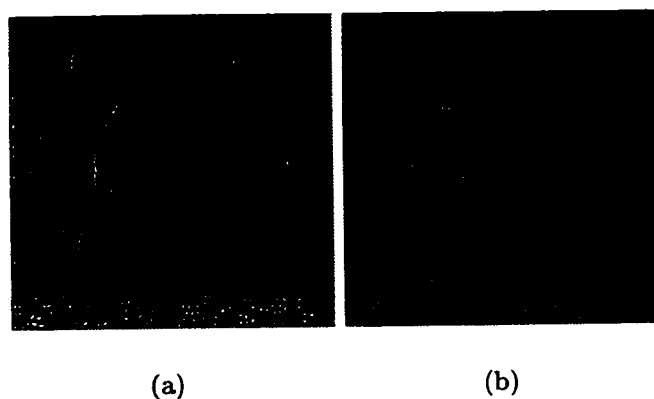
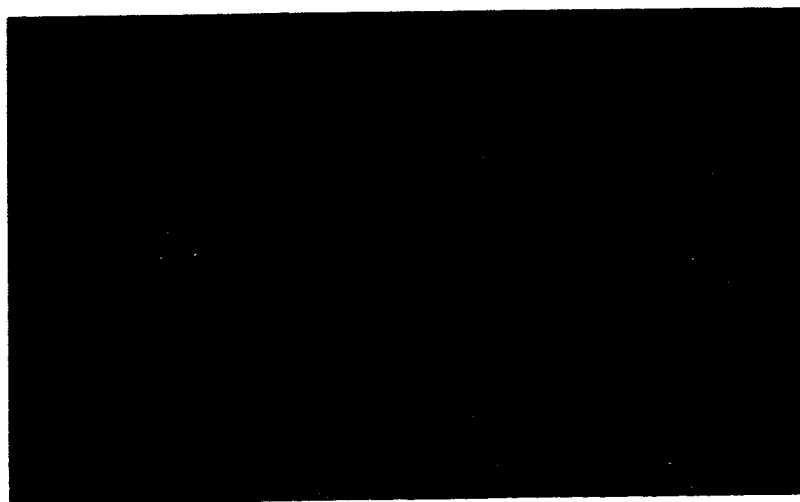
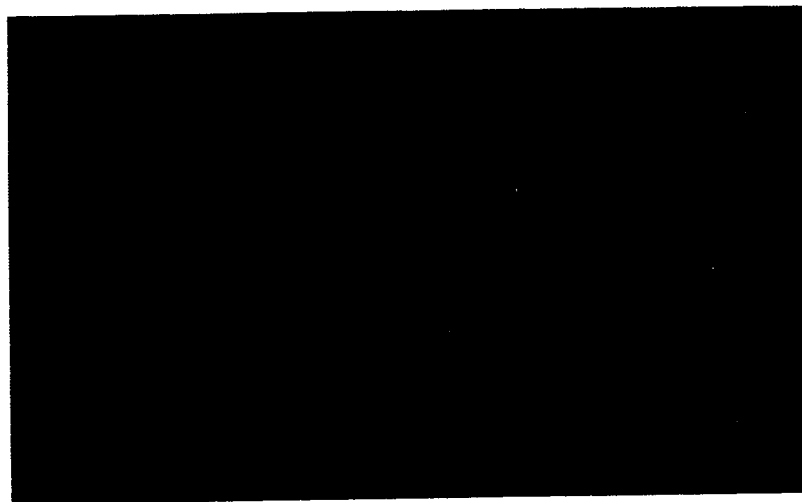


Figure 6.17 : Composantes de luminance passe-haut de la région de fixation de la deuxième image de la séquence *aqua* avec la méthode *intégrée* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

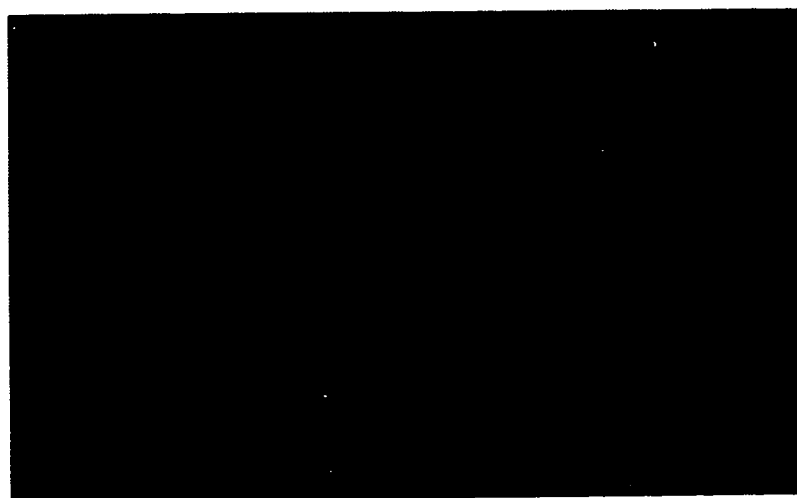


(a)

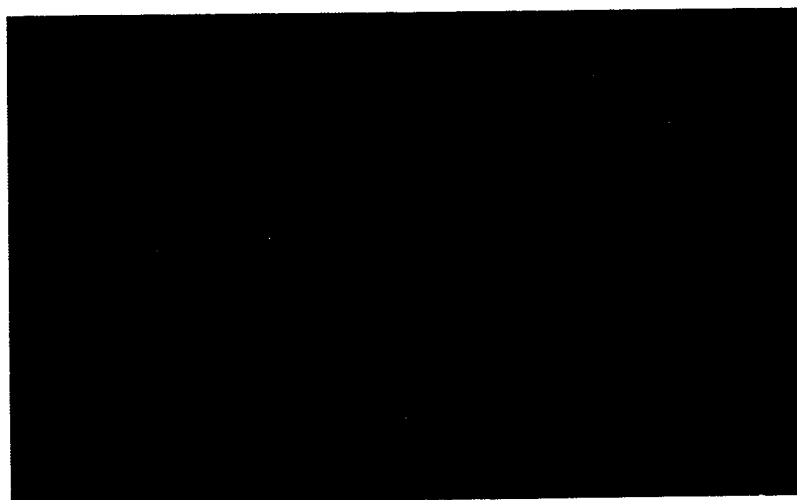


(b)

Figure 6.18 : Composantes de luminance passe-haut agrandies de la sous-région de fixation centrale de la deuxième image de la séquence *aqua* avec la méthode *intégrée* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.



(a)



(b)

Figure 6.19 : Composantes de luminance passe-haut agrandies de la sous-région de fixation adjacente de la deuxième image de la séquence *aqua* avec la méthode *intégrée* : (a) composante originale et (b) composante reconstruite. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

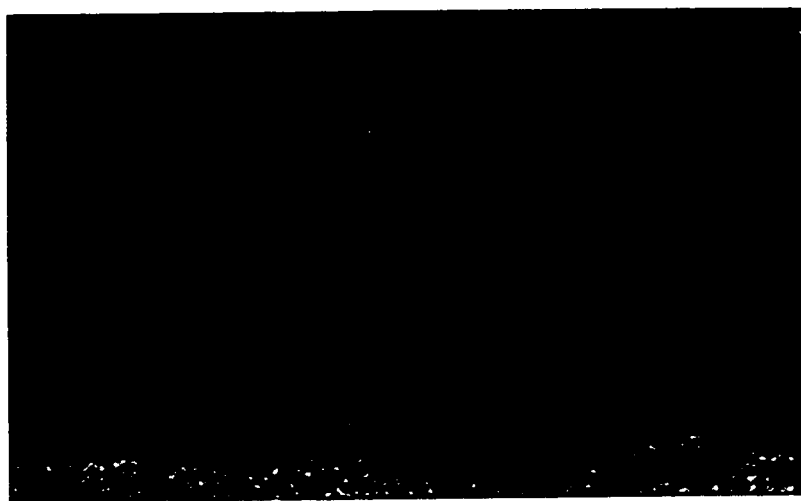


Figure 6.20 : Composante de luminance reconstruite de la deuxième image de la séquence *aqua* avec la méthode *diffusion*.

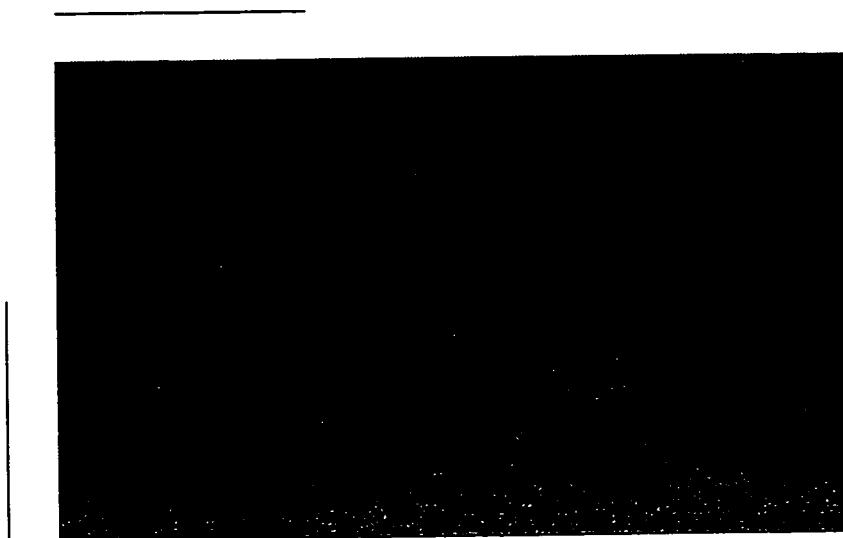


Figure 6.21 : Différence entre les composantes de luminance reconstruite et originale de la deuxième image de la séquence *aqua* avec la méthode *intégrée*. Les lignes horizontale et verticale en bordure de l'image indiquent la région de fixation. Une technique d'amélioration de contraste a été utilisée afin de faciliter la visualisation sur papier.

Tableau 6.1 : Echelle de visibilité de dégradations

critère	pointage
non perceptibles	1
juste perceptibles	2
légères mais définitivement perceptibles	3
non dérangeantes	4
quelque peu dérangeantes	5
définitivement dérangeantes	6
extrêmement dérangeantes	7

par écrit une description des artefacts qui étaient perçus. Deux des trois sujets étaient relativement familiers avec le domaine du traitement d'image, et donc plus attentifs aux différents types de dégradations pouvant se présenter. La procédure, les conditions expérimentales, ainsi que les résultats détaillés sont présentés à l'annexe E.

Les cinq séquences ont été évaluées en prenant comme référence les séquences originales non-traitées. Pour chaque séquence, un attracteur d'attention était présenté afin que les sujets regardent principalement la région de fixation ayant été préalablement choisie. Les principaux artefacts identifiés par les sujets qui sont communs à toutes les méthodes de codage et à toutes les séquences sont des couleurs moins brillantes et des images plus floues. De plus pour certaines séquences, des dédoublements de contour sont quelquefois visibles ainsi qu'une démarcation entre la région de fixation et le fond se traduisant par le papillotement d'une ligne horizontale ou verticale. Néanmoins, la qualité des images codées est satisfaisante, avec une appréciation selon l'échelle du tableau 6.1 se situant en général entre la cote 2 et la cote 4 (inclusive-

Tableau 6.2 : Appréciation moyenne des six méthodes de codage

	appréciation
<i>studio</i>	2,0
<i>studio</i> (spat./temp.)	2,2
<i>diffusion</i>	2,9
<i>diffusion</i> (spat./temp.)	2,9
<i>intégrée</i> (f_l)	2,7
<i>intégrée</i> ($2f_{sp}$)	2,7

ment). Le tableau 6.2 donne une appréciation moyenne des six méthodes de codage. L'appréciation détaillée des trois sujets se trouve à l'annexe E. Cette annexe décrit aussi des artefacts n'ayant pas été systématiquement observés par tous les sujets et qui sont particuliers à certaines séquences ou certaines méthodes de codage. La cause possible de ces artefacts ainsi que quelques suggestions pour les réduire ou les éliminer sont aussi présentées.

Les méthodes *studio* et *studio* avec trous de Fukinuki obtiennent en moyenne la meilleure évaluation. Ceci s'explique par le fait que, par rapport aux autres méthodes, la région de fixation est plus grande, la largeur de bande des composantes du luminance passe-bas est plus élevée et la largeur de bande des composantes de chrominance est la même ou plus élevée que celle des autres méthodes. Les méthodes *diffusion* et *diffusion* avec trous de Fukinuki sont jugées en moyenne équivalentes. De même, les méthodes *intégrées* sont aussi toutes les deux jugées en moyenne équivalentes et perçues comme légèrement supérieures aux méthodes *diffusion*. Toutefois, cet avantage provient probablement de la variance statistique des données puisque les condi-

tions *diffusion* avec trous de Fukinuki et *intégrée* limitée à 4.2 MHz sont à toutes fins pratiques identiques.

6.3 Evaluation objective

L'évaluation objective de la qualité des séquences codées est faite à l'aide de la mesure du rapport signal à bruit crête-à-crête (RSBC) entre les signaux originaux et traités. Cette mesure, donnée par l'équation 6.1, a été choisie car elle est couramment utilisée dans la domaine du codage d'images. Dans cette équation, I et I' représentent respectivement les images originale et reconstruite, toutes les deux de taille $M \times N$. Toutes les valeurs associées à la mesure RSBC sont données en décibels (dB). L'évaluation porte sur les trois principales étapes de traitement nécessaires à la transmission des images, soit la représentation originale des images en composantes à largeur de bande limitée, la modulation et la démodulation des composantes transmises et la reconstruction des images en recombinaison ces composantes. Ces trois aspects sont évalués pour la transmission des composantes de luminance passe-bas qui affectent l'image entière, les composantes de luminance de la région de fixation et les composantes de chrominance.

$$RSBC = 10 \log_{10} \frac{(255)^2}{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |I(m, n) - I'(m, n)|^2}, \quad (6.1)$$

6.3.1 Transmission des composantes de luminance passe-bas

Représentation par composantes Le tableau 6.3 présente le rapport signal à bruit entre les composantes de luminance originales et passe-bas avant la formation du signal vidéo composite. A cette étape, les composantes de luminance sont filtrées pour obtenir des composantes passe-haut et passe-bas, mais ces composantes ne sont pas encore combinées pour former le signal vidéo composite. Les mesures présentées au tableau 6.3 indiquent le meilleur niveau de qualité pouvant être atteint pour les composantes de luminance dans la région périphérique avec la représentation associée aux méthodes de codage proposées. Pour la méthode *studio*, les composantes passe-bas sont filtrées horizontalement à $f_{sp} = 3,58$ MHz, alors qu'une fréquence de coupure horizontale de $f_l/2 = 2,1$ MHz est utilisée pour les méthodes *diffusion* et *intégrée*. Les valeurs RSBC de la table 6.3 dépendent de la quantité de détails des images originales qui sont supprimées par le filtrage passe-bas. Des différences allant jusqu'à 2,6 dB sont observées entre les canaux gauche et droit d'une même séquence. Ces différences peuvent être expliquées par des variations de la géométrie de la scène, des conditions d'illumination et de la sensibilité des caméras entre les deux points de vue. La différence entre les valeurs RSBC des deux canaux d'une même séquence demeure relativement constante durant les différentes étapes de traitement des différentes méthodes de codage et sera donc aussi observée dans les mesures RSBC reportées subséquemment. Des différences allant de 1,2 à 2,3 dB sont observées entre la méthode *studio* et les autres méthodes puisqu'un filtre avec une fréquence de coupure plus élevée est utilisé.

Tableau 6.3 : RSBC entre les composantes de luminance originales et passe-bas avant la formation du signal vidéo composite.

	Y ga.	Y dr.
Tunnel (f_{sp})	28,73	28,62
Piano (f_{sp})	28,13	30,71
Aqua (f_{sp})	29,84	27,77
Manège (f_{sp})	28,21	29,92
Train (f_{sp})	32,86	32,72
Tunnel ($f_l/2$)	27,50	27,36
Piano ($f_l/2$)	25,80	28,37
Aqua ($f_l/2$)	27,77	25,82
Manège ($f_l/2$)	26,38	28,03
Train ($f_l/2$)	31,84	31,57

Modulation et démodulation Le tableau 6.4 présente le rapport signal à bruit entre les composantes de luminance passe-bas originales et reconstruites. Pour les méthodes *studio* et *diffusion*, les résultats sont donnés pour la reconstruction avec les filtrages spatial et spatio-temporel des composantes de chrominance. Les mesures présentées au tableau 6.3 indiquent l'efficacité de la transmission des composantes de luminance passe-bas. Les valeurs du tableau 6.4 sont effectivement assez élevées, indiquant une transmission efficace des composantes de luminance passe-bas. Les valeurs RSBC pour la composante passe-bas droite transmise avec les méthodes *diffusion* et *intégrée* limitée à 4,2 MHz sont moins élevées puisque cette composante est modulée avec bande latérale résiduelle (VSB), plutôt qu'avec double bande latérale pour les autres méthodes. Pour les méthodes *studio* et *diffusion*, le choix de la méthode de reconstruction des composantes de chrominance n'affecte pas de façon significative la

Tableau 6.4 : RSBC entre les composantes de luminance passe-bas originales et passe-bas reconstruites.

	Y ga. (spat./temp.)	Y dr. (spat./temp.)
Tunnel <i>studio</i>	39,30/39,30	39,39/39,34
Piano <i>studio</i>	38,78/38,75	39,64/39,65
Aqua <i>studio</i>	41,22/41,23	40,13/40,15
Manège <i>studio</i>	41,62/41,58	42,78/42,82
Train <i>studio</i>	43,66/43,61	43,52/43,61
Tunnel <i>diffusion</i>	40,45/40,45	38,85/38,85
Piano <i>diffusion</i>	38,75/38,78	38,18/38,11
Aqua <i>diffusion</i>	39,33/39,34	37,49/37,43
Manège <i>diffusion</i>	39,40/39,42	38,94/38,75
Train <i>diffusion</i>	44,80/44,81	42,25/42,20
Tunnel <i>intégrée</i> ($2f_{sp}$)	41,32	40,50
Piano <i>intégrée</i> ($2f_{sp}$)	39,74	39,37
Aqua <i>intégrée</i> ($2f_{sp}$)	40,82	39,00
Manège <i>intégrée</i> ($2f_{sp}$)	40,44	40,26
Train <i>intégrée</i> ($2f_{sp}$)	45,72	44,05
Tunnel <i>intégrée</i> (f_l)	41,33	38,43
Piano <i>intégrée</i> (f_l)	39,72	36,78
Aqua <i>intégrée</i> (f_l)	40,83	36,94
Manège <i>intégrée</i> (f_l)	40,39	37,68
Train <i>intégrée</i> (f_l)	45,81	41,51

qualité de transmission des composantes passe-bas.

Reconstruction Le tableau 6.5 présente le rapport signal à bruit entre les composantes de luminance originales et passe-bas reconstruites. Pour les méthodes *studio* et *diffusion*, les résultats sont donnés pour la reconstruction avec les filtrages spatial et spatio-temporel des composantes de chrominance. Les mesures présentées au tableau 6.5 indiquent la qualité des composantes de luminance passe-bas recon-

struites par rapport aux composantes de de luminance originales. Ces mesures sont une indication de la performance des méthodes de compression et incorpore les effets introduits par la représentation par composante, la formation du signal vidéo composite et la reconstruction des composantes de luminance passe-bas à partir du signal vidéo composite. Les valeurs du tableau 6.5 représentent le niveau de qualité du signal de luminance reconstruit dans la région périphérique. Une perte de 0,7 à 1,5 dB est observée entre les valeurs du tableau 6.5 et celles de la représentation originale du tableau 6.3. Cette perte représente la dégradation du niveau de qualité du signal causé par la formation du signal vidéo composite et la reconstruction des composantes de luminance passe-bas à partir de ce signal. Les valeurs du tableau 6.5 sont plus élevées pour la méthode *studio* puisqu'un filtre avec une fréquence de coupure plus élevée est utilisé. Pour les méthodes *studio* et *diffusion*, le choix de la méthode de reconstruction des composantes de chrominance n'affecte pas de façon significative la qualité de transmission des composantes passe-bas.

6.3.2 Transmission des composantes de luminance de la région de fixation

Représentation par composantes Le tableau 6.6 présente le rapport signal à bruit entre les composantes de luminance originales et passe-bas de la région de fixation avant la formation du signal vidéo composite. Pour la méthode *intégrée*, les résultats sont donnés pour des largeurs de bande de 4,2 MHz et de 7,16 MHz

Tableau 6.5 : RSBC entre les composantes de luminance originales et passe-bas re-construites.

	Y ga. (spat./temp.)	Y dr. (spat./temp.)
Tunnel <i>studio</i>	27,35/27,35	27,31/27,31
Piano <i>studio</i>	26,87/26,87	29,24/29,24
Aqua <i>studio</i>	28,65/28,65	26,75/26,75
Manège <i>studio</i>	27,24/27,24	28,94/28,94
Train <i>studio</i>	31,65/31,64	31,56/31,57
Tunnel <i>diffusion</i>	26,67/26,67	26,43/26,43
Piano <i>diffusion</i>	25,04/25,04	27,14/27,13
Aqua <i>diffusion</i>	26,80/26,80	24,83/24,83
Manège <i>diffusion</i>	25,56/25,56	26,95/26,94
Train <i>diffusion</i>	31,00/31,00	30,64/30,64
Tunnel <i>intégrée</i> ($2f_{sp}$)	26,70	26,51
Piano <i>intégrée</i> ($2f_{sp}$)	25,08	27,26
Aqua <i>intégrée</i> ($2f_{sp}$)	26,87	24,94
Manège <i>intégrée</i> ($2f_{sp}$)	25,60	27,09
Train <i>intégrée</i> ($2f_{sp}$)	31,03	30,63
Tunnel <i>intégrée</i> (f_l)	26,70	26,41
Piano <i>intégrée</i> (f_l)	25,07	27,01
Aqua <i>intégrée</i> (f_l)	26,87	24,80
Manège <i>intégrée</i> (f_l)	25,59	26,85
Train <i>intégrée</i> (f_l)	31,03	30,59

puisque la taille de région de fixation est différente pour chaque cas. Les résultats pour la méthode diffusion sont les mêmes que ceux pour la méthode *intégrée* limitée à 4,2 MHz. A la section 6.3.1, des mesures similaires ont été présentées pour les composantes de luminance affectant l'image complète. Les données du tableau 6.6 sont présentées afin de mieux quantifier l'amélioration de la qualité du signal causée par la transmission d'information de hautes-fréquences pour la région de fixation. En effet, les valeurs du tableau 6.6 ne sont pas toujours similaires à celles du tableau 6.3 puisqu'elles ne concernent pas les mêmes régions de l'image. Une différence marquée est observée pour la séquence *piano* où les valeurs du tableau 6.6 sont nettement inférieures à celles du tableau 6.3. Ceci indique qu'une grande quantité de détails sont supprimés par le filtrage passe-bas dans cette région, ce qui n'est pas tellement surprenant puisque la région de fixation pour cette séquence a effectivement été choisie dans le but de tester les méthodes de codage sur une région de fixation contenant beaucoup de détails.

Le tableau 6.6 présente le rapport signal à bruit entre les composantes de luminance originales et la somme des composantes passe-bas et passe-haut de la région de fixation avant la formation du signal vidéo composite. Les résultats pour la méthode *diffusion* sont les mêmes que ceux pour la méthode *intégrée* limitée à 4,2 MHz. Les mesures présentées au tableau 6.3 indiquent le meilleur niveau de qualité pouvant être atteint pour les composantes de luminance dans la région de fixation avec la représentation associée aux méthodes de codage proposées. La représentation con-

Tableau 6.6 : RSBC entre les composantes de luminance originales et passe-bas de la région de fixation avant la formation du signal vidéo composite.

	$Y_{ga.} (f_l, 2f_{sp})$	$Y_{dr.} (f_l, 2f_{sp})$
Tunnel (f_{sp})	30,39	30,71
Piano (f_{sp})	24,33	27,26
Aqua (f_{sp})	29,66	27,16
Manège (f_{sp})	28,60	28,69
Train (f_{sp})	31,87	31,28
Tunnel ($f_l/2$)	27,46/28,49	27,50/28,30
Piano ($f_l/2$)	20,43/20,95	23,33/25,38
Aqua ($f_l/2$)	27,56/27,54	25,17/24,87
Manège ($f_l/2$)	25,62/26,80	27,20/27,16
Train ($f_l/2$)	28,98/30,54	28,75/29,98

sistant en une composante passe-bas et une composante passe-haut n'est pas explicitement calculée pour la région de fixation puisque ce sont les composantes passe-haut agrandies qui sont impliquées dans la formation du signal vidéo composite.

Un traitement additionnel consistant à réduire les composantes passe-haut agrandies a donc été accompli afin d'obtenir les mesures du tableau 6.7. Une amélioration du niveau de qualité des signaux de luminance de 11 à 13 dB est obtenu par rapport aux composantes passe-bas du tableau 6.6.

Modulation et démodulation Le tableau 6.8 présente le rapport signal à bruit entre les composantes de luminance passe-bas originales et reconstruites de la région de fixation. Pour les méthodes *studio* et *diffusion*, les résultats sont donnés pour la reconstruction avec les filtrages spatial et spatio-temporel des composantes de chrominance. Les mesures présentées au tableau 6.8 indiquent l'efficacité de la trans-

Tableau 6.7 : RSBC entre les composantes de luminance originales et la somme des composantes passe-bas et passe-haut de la région de fixation avant la formation du signal vidéo composite.

	$Y_{ga.}$	$Y_{dr.}$
<i>Tunnel studio</i>	41,91	41,93
<i>Piano studio</i>	35,29	38,39
<i>Aqua studio</i>	41,60	39,09
<i>Manège studio</i>	39,77	39,87
<i>Train studio</i>	43,14	42,49
<i>Tunnel intégrée ($2f_{sp}$)</i>	40,55	40,57
<i>Piano intégrée ($2f_{sp}$)</i>	33,03	37,52
<i>Aqua intégrée ($2f_{sp}$)</i>	39,76	36,84
<i>Manège intégrée ($2f_{sp}$)</i>	39,35	39,53
<i>Train intégrée ($2f_{sp}$)</i>	43,50	42,51
<i>Tunnel intégrée (f_l)</i>	39,72	39,74
<i>Piano intégrée (f_l)</i>	32,25	35,42
<i>Aqua intégrée (f_l)</i>	39,73	36,81
<i>Manège intégrée (f_l)</i>	37,75	38,94
<i>Train intégrée (f_l)</i>	42,34	41,80

mission des composantes de luminance passe-bas de la région de fixation. Comme à la section 6.3.1, les valeurs du tableau 6.8 sont élevées, indiquant une transmission efficace des composantes de luminance passe-bas. Les valeurs RSBC pour la composante passe-bas droite transmise avec les méthodes *diffusion* et *intégrée* limitée à 4,2 MHz sont moins élevées puisque cette composante est modulée avec bande latérale résiduelle (VSB), plutôt qu'avec double bande latérale pour les autres méthodes. Pour les méthodes *studio* et *diffusion*, le choix de la méthode de reconstruction des composantes de chrominance n'affecte pas de façon significative la qualité de transmission des composantes passe-bas.

Le tableau 6.9 présente le rapport signal à bruit entre la somme des composantes de luminance passe-bas et passe-haut originales de la région de fixation et la somme des composantes passe-bas et passe-haut reconstruites de la région de fixation. Les mesures présentées au tableau 6.9 indiquent l'efficacité de la transmission de l'information de luminance correspondant à la région de fixation. Il n'est pas espéré d'obtenir des valeurs RSBC aussi élevées que celles du tableau 6.8 parce que plus d'opérations sont accomplies sur les composantes de luminance passe-haut, telles que du filtrage, de l'interpolation, de la décimation, de la modulation et de la démodulation. Les dégradations introduites par chacune de ces opérations sont cumulatives et affectent la qualité de reconstruction des signaux transmis. Comme pour les valeurs du tableau 6.7, l'obtention de celles du tableau 6.9 nécessite un traitement additionnel à l'émetteur pour obtenir la représentation consistant en une composante passe-bas et une com-

Tableau 6.8 : RSBC entre les composantes de luminance passe-bas originales et passe-bas reconstruites de la région de fixation.

	Y ga. (spat./temp.)	Y dr. (spat./temp.)
Tunnel <i>studio</i>	41,39/41,39	40,60/40,54
Piano <i>studio</i>	36,10/36,10	36,87/36,85
Aqua <i>studio</i>	42,35/42,36	39,50/39,50
Manège <i>studio</i>	42,21/42,24	42,10/42,15
Train <i>studio</i>	45,02/45,02	44,82/44,88
Tunnel <i>diffusion</i>	40,62/40,62	39,22/39,25
Piano <i>diffusion</i>	34,50/34,50	33,66/33,67
Aqua <i>diffusion</i>	39,45/39,46	37,67/37,60
Manège <i>diffusion</i>	39,85/39,86	39,42/39,24
Train <i>diffusion</i>	42,99/42,99	41,62/41,67
Tunnel <i>intégrée</i> ($2f_{sp}$)	43,74	41,50
Piano <i>intégrée</i> ($2f_{sp}$)	37,11	37,07
Aqua <i>intégrée</i> ($2f_{sp}$)	42,30	38,77
Manège <i>intégrée</i> ($2f_{sp}$)	42,63	40,85
Train <i>intégrée</i> ($2f_{sp}$)	45,94	43,71
Tunnel <i>intégrée</i> (f_l)	41,79	38,66
Piano <i>intégrée</i> (f_l)	35,70	32,57
Aqua <i>intégrée</i> (f_l)	40,09	37,05
Manège <i>intégrée</i> (f_l)	40,47	38,76
Train <i>intégrée</i> (f_l)	43,83	41,18

Tableau 6.9 : RSBC entre la somme des composantes de luminance passe-bas et passe-haut originales de la région de fixation et la somme des composantes passe-bas et passe-haut reconstruites de la région de fixation.

	Y ga.	Y dr.
Tunnel <i>studio</i>	33,43	33,84
Piano <i>studio</i>	29,12	30,76
Aqua <i>studio</i>	35,54	33,88
Manège <i>studio</i>	33,56	33,80
Train <i>studio</i>	35,10	34,41
Tunnel <i>diffusion</i>	32,05	32,12
Piano <i>diffusion</i>	26,31	28,02
Aqua <i>diffusion</i>	31,73	30,19
Manège <i>diffusion</i>	30,03	31,29
Train <i>diffusion</i>	33,31	32,95
Tunnel <i>intégrée</i> ($2f_{sp}$)	33,04	32,43
Piano <i>intégrée</i> ($2f_{sp}$)	26,38	29,62
Aqua <i>intégrée</i> ($2f_{sp}$)	31,85	29,57
Manège <i>intégrée</i> ($2f_{sp}$)	31,18	31,36
Train <i>intégrée</i> ($2f_{sp}$)	34,75	34,18
Tunnel <i>intégrée</i> (f_l)	32,11	31,33
Piano <i>intégrée</i> (f_l)	26,34	26,97
Aqua <i>intégrée</i> (f_l)	31,96	29,35
Manège <i>intégrée</i> (f_l)	30,06	30,68
Train <i>intégrée</i> (f_l)	33,40	32,45

posante passe-haut pour la région de fixation. Les valeurs du tableau 6.9 sont relativement élevées, indiquant une transmission satisfaisante des composantes de luminance de la région de fixation.

Reconstruction Le tableau 6.10 présente le rapport signal à bruit entre les composantes de luminance originales et passe-bas reconstruites de la région de fixation. Pour les méthodes *studio* et *diffusion*, les résultats sont donnés pour la reconstruc-

tion avec les filtrages spatial et spatio-temporel des composantes de chrominance. Une perte de 0,5 à 1,4 dB est observée entre les valeurs du tableau 6.10 et celles de la représentation originale du tableau 6.6. Cette perte représente la dégradation du niveau de qualité du signal causé par la formation du signal vidéo composite et la reconstruction des composantes de luminance passe-bas à partir de ce signal. Les valeurs du tableau 6.10 sont plus élevées pour la méthode *studio* puisqu'un filtre avec une fréquence de coupure plus élevée est utilisé. Pour les méthodes *studio* et *diffusion*, le choix de la méthode de reconstruction des composantes de chrominance n'affecte pas de façon significative la qualité de transmission des composantes passe-bas.

Le tableau 6.11 présente le rapport signal à bruit entre les composantes de luminance originales et reconstruites de la région de fixation. Pour les méthodes *studio* et *diffusion*, les résultats sont donnés pour la reconstruction avec les filtrages spatial et spatio-temporel des composantes de chrominance. Les valeurs du tableau 6.11 varient de 27 à 35 dB pour la méthode *studio*, et entre 23,5 et 33 dB pour les autres méthodes. La meilleure performance de la méthode *studio* peut être expliquée par le fait que moins d'opérations de filtrage, de modulation et de démodulation sont accomplies avec cette méthode. Une augmentation du niveau de qualité du signal de luminance de la région de fixation de 2,5 à 5,9 dB est observée lorsque l'information de hautes-fréquences est ajoutée aux composantes passe-bas du tableau 6.10. Les valeurs relativement peu élevées obtenues pour la séquence *piano* avec les méthodes *diffusion* et intégrée limitée à 4,2 MHz s'explique par le fait que les composantes de luminance

Tableau 6.10 : RSBC entre les composantes de luminance originales et passe-bas reconstruites de la région de fixation.

	Y ga. (spat./temp.)	Y dr. (spat./temp.)
Tunnel <i>studio</i>	29,04/29,04	29,41/29,41
Piano <i>studio</i>	23,21/23,21	25,95/25,94
Aqua <i>studio</i>	28,60/28,60	26,19/26,19
Manège <i>studio</i>	27,59/27,59	27,70/27,70
Train <i>studio</i>	30,77/30,77	30,27/30,27
Tunnel <i>diffusion</i>	26,68/26,68	26,52/26,52
Piano <i>diffusion</i>	19,80/19,80	22,15/22,15
Aqua <i>diffusion</i>	26,63/26,63	24,28/24,28
Manège <i>diffusion</i>	24,88/24,88	26,30/26,29
Train <i>diffusion</i>	28,26/28,26	27,91/27,91
Tunnel <i>intégrée</i> ($2f_{sp}$)	27,87	27,41
Piano <i>intégrée</i> ($2f_{sp}$)	20,45	24,36
Aqua <i>intégrée</i> ($2f_{sp}$)	26,89	24,07
Manège <i>intégrée</i> ($2f_{sp}$)	26,21	26,37
Train <i>intégrée</i> ($2f_{sp}$)	29,93	29,19
Tunnel <i>intégrée</i> (f_l)	26,71	26,49
Piano <i>intégrée</i> (f_l)	19,84	22,07
Aqua <i>intégrée</i> (f_l)	26,72	24,24
Manège <i>intégrée</i> (f_l)	24,90	26,26
Train <i>intégrée</i> (f_l)	28,29	27,89

Tableau 6.11 : RSBC entre les composantes de luminance originales et reconstruites de la région de fixation.

	Y ga. (spat./temp.)	Y dr. (spat./temp.)
Tunnel <i>studio</i>	32,01/32,01	32,34/32,35
Piano <i>studio</i>	27,12/27,16	29,17/29,28
Aqua <i>studio</i>	33,95/33,93	32,05/32,01
Manège <i>studio</i>	31,71/31,68	31,89/31,90
Train <i>studio</i>	34,65/34,62	34,30/34,30
Tunnel <i>diffusion</i>	29,91/29,88	29,93/29,55
Piano <i>diffusion</i>	23,69/23,63	25,83/25,38
Aqua <i>diffusion</i>	29,72/29,70	27,72/27,34
Manège <i>diffusion</i>	28,13/28,12	29,44/29,21
Train <i>diffusion</i>	31,60/31,63	31,20/31,10
Tunnel <i>intégrée</i> ($2f_{sp}$)	30,87	30,47
Piano <i>intégrée</i> ($2f_{sp}$)	23,95	27,61
Aqua <i>intégrée</i> ($2f_{sp}$)	29,77	27,30
Manège <i>intégrée</i> ($2f_{sp}$)	29,36	29,53
Train <i>intégrée</i> ($2f_{sp}$)	33,01	32,28
Tunnel <i>intégrée</i> (f_l)	29,93	29,38
Piano <i>intégrée</i> (f_l)	23,70	25,09
Aqua <i>intégrée</i> (f_l)	29,86	27,16
Manège <i>intégrée</i> (f_l)	28,14	28,98
Train <i>intégrée</i> (f_l)	31,66	30,82

dans la région de fixation possèdent de faibles ratios signal à bruit : 20,43 et 23,33 dB pour les composantes passe-bas originales et 19,80 et 22,15 dB pour les composantes de luminance passe-bas reconstruites. Les valeurs du tableau 6.5 de 23,69 et 25,83 dB pour la séquence *piano* représentent donc une amélioration du niveau de qualité du signal de 3,89 et 3,68 dB, ce qui est comparable aux autres séquences.

6.3.3 Transmission des composantes de chrominance

Tableau 6.12 : RSBC entre les composantes de chrominance originales et passe-bas avant la formation du signal vidéo composite.

	I gauche	I droit	Q gauche	Q droit
Tunnel ($3f_{sp}/7, f_{sp}/7$)	36,41	36,15	35,06	37,79
Piano ($3f_{sp}/7, f_{sp}/7$)	37,76	38,77	35,35	37,92
Aqua ($3f_{sp}/7, f_{sp}/7$)	36,94	35,83	35,54	36,39
Manège ($3f_{sp}/7, f_{sp}/7$)	36,32	37,94	37,67	40,07
Train ($3f_{sp}/7, f_{sp}/7$)	37,76	38,76	35,35	37,92
Tunnel ($f_{sp}/7, f_{sp}/7$)	34,42	34,17	35,06	37,79
Piano ($f_{sp}/7, f_{sp}/7$)	35,23	36,23	35,35	37,92
Aqua ($f_{sp}/7, f_{sp}/7$)	35,09	34,20	35,54	36,39
Manège ($f_{sp}/7, f_{sp}/7$)	33,96	35,54	37,67	40,07
Train ($f_{sp}/7, f_{sp}/7$)	35,23	36,23	35,35	37,92

Représentation par composantes Le tableau 6.12 présente le rapport signal à bruit entre les composantes de chrominance originales et passe-bas avant la formation du signal vidéo composite. Dans ce tableau, les chiffres entre parenthèses indiquent la largeur de bande des composantes Q et I . Les mesures présentées au tableau 6.12 indiquent le meilleur niveau de qualité pouvant être atteint pour les composantes de chrominance avec la représentation associée aux méthodes de codage proposées. Une dégradation du niveau de qualité de la composante I est observée lorsque celle-ci est limitée à $f_{sp}/7 = 0,5$ MHz. Les valeurs du tableau 6.12 pour la composante Q sont systématiquement plus élevées pour le canal droit que pour celui de gauche, présumément en raison d'une différence de sensibilité entre les caméras des deux canaux.

Modulation et démodulation Le tableau 6.13 présente le rapport signal à bruit entre les composantes de chrominance passe-bas originales et passe-bas reconstruites. Dans ce tableau, les chiffres entre parenthèses indiquent la largeur de bande des composantes Q et I , *temp.* indique une modulation temporelle en quadrature avec double bande latérale et *temp. rés.* une modulation en quadrature avec bande latérale résiduelle (VSB). Les mesures présentées au tableau 6.13 indiquent l'efficacité de la transmission des composantes de chrominance passe-bas. Les valeurs du tableau 6.13 sont élevées, indiquant une transmission efficace des composantes de chrominance. Les valeurs plus élevées pour la composante I à $f_{sp}/7 = 0,5$ MHz s'explique par le fait que la mesure du rapport signal à bruit se base sur une composante de référence contenant moins de détails que pour les autres méthodes.

Reconstruction Le tableau 6.14 présente le rapport signal à bruit entre les composantes de chrominance originales et passe-bas reconstruites. Dans ce tableau, les chiffres entre parenthèses indiquent la largeur de bande des composantes Q et I , *temp.* indique une modulation temporelle en quadrature avec double bande latérale et *temp. rés.* une modulation en quadrature avec bande latérale résiduelle. Les valeurs du tableau 6.14 pour la composante I varient de 33,5 à 36 dB lorsque la largeur de bande est de $3f_{sp}/7 = 1,5$ MHz et de 32 à 34 dB lorsqu'elle est limitée à $f_{sp}/7 = 0,5$ MHz. Les valeurs pour la composante Q varient de 34 à 38,5 dB. Pour la modulation temporelle des composantes de chrominance afin d'exploiter les trous de Fukinuki, l'utilisation d'une double bande latérale ou d'une bande latérale résiduelle est presque équivalente,

Tableau 6.13 : RSBC entre les composantes de chrominance passe-bas originales et passe-bas reconstruites.

	I gauche	I droit	Q gauche	Q droit
Tunnel ($3f_{sp}/7, f_{sp}/7$)	38,32	38,23	44,41	44,82
Piano ($3f_{sp}/7, f_{sp}/7$)	37,74	37,73	43,03	43,21
Aqua ($3f_{sp}/7, f_{sp}/7$)	39,38	39,30	44,00	44,04
Manège ($3f_{sp}/7, f_{sp}/7$)	38,37	38,70	46,97	48,15
Train ($3f_{sp}/7, f_{sp}/7$)	40,39	40,10	46,86	47,37
Tunnel ($3f_{sp}/7, f_{sp}/7$ temp.)	40,73	39,80	47,86	47,02
Piano ($3f_{sp}/7, f_{sp}/7$ temp.)	39,91	40,07	42,18	41,87
Aqua ($3f_{sp}/7, f_{sp}/7$ temp.)	41,58	42,20	42,70	43,73
Manège ($3f_{sp}/7, f_{sp}/7$ temp.)	37,34	37,84	43,80	44,60
Train ($3f_{sp}/7, f_{sp}/7$ temp.)	39,19	39,78	46,37	46,49
Tunnel ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	40,82	40,04	48,27	47,52
Piano ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	40,44	40,09	43,09	41,96
Aqua ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	40,42	42,95	41,24	44,82
Manège ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	37,40	37,90	44,12	44,88
Train ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	39,28	39,70	46,87	46,12
Tunnel ($f_{sp}/7, f_{sp}/7$)	41,01	41,04	44,59	45,96
Piano ($f_{sp}/7, f_{sp}/7$)	40,07	41,05	42,37	44,77
Aqua ($f_{sp}/7, f_{sp}/7$)	44,01	44,11	44,32	44,86
Manège ($f_{sp}/7, f_{sp}/7$)	41,69	43,90	46,93	48,72
Train ($f_{sp}/7, f_{sp}/7$)	41,49	41,44	45,69	46,90

Tableau 6.14 : RSBC entre les composantes de chrominance originales et passe-bas reconstruites.

	I gauche	I droit	Q gauche	Q droit
Tunnel ($3f_{sp}/7, f_{sp}/7$)	33,67	33,46	33,98	36,20
Piano ($3f_{sp}/7, f_{sp}/7$)	34,27	34,78	34,13	36,14
Aqua ($3f_{sp}/7, f_{sp}/7$)	34,51	33,76	34,26	34,92
Manège ($3f_{sp}/7, f_{sp}/7$)	33,69	34,87	36,50	38,63
Train ($3f_{sp}/7, f_{sp}/7$)	34,78	34,17	36,46	38,28
Tunnel ($3f_{sp}/7, f_{sp}/7$ temp.)	34,57	34,16	34,64	37,14
Piano ($3f_{sp}/7, f_{sp}/7$ temp.)	35,24	36,05	34,39	36,44
Aqua ($3f_{sp}/7, f_{sp}/7$ temp.)	35,30	34,60	34,63	35,53
Manège ($3f_{sp}/7, f_{sp}/7$ temp.)	33,47	34,60	36,38	38,30
Train ($3f_{sp}/7, f_{sp}/7$ temp.)	34,57	34,43	36,74	38,49
Tunnel ($3f_{sp}/7, f_{sp}/7$ temp. rés.)	34,59	34,23	34,66	37,18
Piano ($3f_{sp}/7, f_{sp}/7$ chrominance. rés.)	35,44	36,08	34,52	36,48
Aqua ($3f_{sp}/7, f_{sp}/7$ chrominance. rés.)	34,98	34,71	34,37	35,67
Manège ($3f_{sp}/7, f_{sp}/7$ chrominance. rés.)	33,50	34,64	36,43	38,37
Train ($3f_{sp}/7, f_{sp}/7$ chrominance. rés.)	34,61	34,40	36,78	38,43
Tunnel ($f_{sp}/7, f_{sp}/7$)	32,46	32,26	34,01	36,36
Piano ($f_{sp}/7, f_{sp}/7$)	32,92	33,94	34,05	36,44
Aqua ($f_{sp}/7, f_{sp}/7$)	33,87	33,18	34,31	35,02
Manège ($f_{sp}/7, f_{sp}/7$)	32,36	34,08	36,51	38,70
Train ($f_{sp}/7, f_{sp}/7$)	33,23	32,81	36,36	38,17

avec des différences entre les deux conditions généralement de l'ordre de 0,1 dB et une différence maximale de 0,4 dB pour la séquence *aqua*. Selon la séquence codée, la modulation temporelle est soit légèrement meilleure (par de 0,25 à 1,3 dB pour I et de 0,2 à 0,9 dB pour Q) ou un peu moins efficace (par de 0,1 à 0,3 dB pour I et Q) que la modulation spatiale.

6.4 Discussion

Les résultats expérimentaux sont satisfaisants et représentatifs des compromis qui existent entre les différentes méthodes de codage. Toutes les méthodes donnent lieu à une amélioration du niveau de qualité dans le région de fixation, tel que désiré. Le niveau de qualité des images codées avec les méthodes *studio* et *studio* avec trous de Fukinuki est en général meilleure qu'avec les autres méthodes en raison, entres autres, d'une plus grande région de fixation, d'une largeur de bande plus élevée pour les composantes de luminance passe-bas et d'une largeur de bande des composantes de chrominance plus élevée ou égale à celle des autres méthodes. De plus, cette méthode est plus simple car elle implique moins de manipulations spectrales que les autres méthodes. La méthode *intégrée* a été développée afin d'être appropriée à la fois pour la largeur de bande du format *studio* et celle du format *diffusion*. Le prix à payer pour cette flexibilité est une région de fixation dont la taille est plus petite qu'avec la méthode *studio* ($2L/5$ plutôt que $L/2$) et des composantes de luminance passe-bas avec une largeur de bande réduite (2.1 MHz plutôt que 3.58 MHz). La largeur de bande du format *diffusion* impose une réduction additionnelle de la taille de la région de fixation qui se traduit par un niveau de qualité des images moins élevé qu'avec les méthodes *studio*. Une transmission des composantes de chrominance en exploitant les trous de Fukinuki doit être favorisée car elle permet l'utilisation d'une plus grande largeur de bande pour ces composantes. Par contre, s'il est impossible d'accomplir un filtrage temporel au récepteur, l'approche de modulation des com-

posantes de chrominance pour le format *studio* s'avère aussi adéquate. Par contre l'approche de modulation de chrominance suggérée avec la méthode *diffusion* qui consiste à limiter les composantes I à 0,5 MHz ne devrait être envisagée qu'en dernier recours, puisque la faible résolution de ces composantes peut dans certaines situations masquer les détails de luminance de la région de fixation. Il est néanmoins important d'observer qu'une limitation de toutes les composantes de chrominance à 0,5 MHz produit des images dont le niveau de qualité est équivalent à celui d'images pouvant être observées sur plusieurs récepteurs de télévision domestique.

L'évaluation subjective des images codées a mis en évidence certains artefacts introduits par les méthodes proposées. Les deux principaux types de dégradations, un ternissement des couleurs et une augmentation du flou des images, sont inhérents à l'approche de codage qui a été choisie. Le ternissement des couleurs provient principalement des limites sur la largeur de bande des composantes de chrominance spécifiée par le standard NTSC, alors que l'augmentation du flou des images provient en partie de notre incapacité de s'assurer que les sujets regardent toujours dans la région de fixation choisie, et en partie des dégradations introduites par les opérations de filtrage et de manipulation spectrale qui font que les images reconstruites ne possèdent plus le même niveau de qualité que les images originales. D'autres types de dégradations de moindre importance ont aussi été identifiées et des solutions possibles pour les éliminer sont présentées à l'annexe E.

Conclusion

Une approche originale de compression spectrale de séquences stéréoscopiques qui exploite la présence d'une région de fixation pour réduire la quantité d'information à transmettre a été présentée. Puisque les acuités spatiales et stéréoscopiques du système visuel humain sont plus élevées autour du point de fixation des yeux et diminuent rapidement avec une augmentation de l'excentricité rétinienne, l'information de haute-fréquences n'est pas transmise pour tout le champ visuel, mais seulement à l'intérieur d'un sous-espace contenant la région de fixation. Cette approche de compression soulève deux problèmes fondamentaux : l'estimation d'une région de fixation et la modulation de la fidélité du codage en fonction de cette région. Ces deux problèmes ont été abordés dans ce travail.

Dans le but de développer une stratégie d'estimation de région de fixation, nous avons accompli une étude psychophysique portant sur les stratégies visuelles utilisées lors de l'accomplissement d'une tâche dynamique de discrimination de profondeur simulée à l'aide de stéréogrammes dynamiques. Cette étude portait sur deux aspects de la performance des sujets lors de l'accomplissement de la tâche choisie : les mouve-

ments de leurs yeux en fonction de la présence de distracteurs et l'effet des conditions particulières associées au visionnement stéréoscopique. Les résultats de l'étude montrent que les fixations des sujets se trouvent majoritairement dans la région centrale de l'écran correspondant à la tâche principale et dans une zone située à droite de celle-ci, ce qui indique que la tâche étudiée peut être accomplie sans déplacement significatif de la région de fixation. Une explication reliée à la géométrie des systèmes de visualisation stéréoscopique a été proposée pour tenir compte des fixations se trouvant à droite de la région centrale. En fonction des résultats de cette étude, une stratégie d'estimation de région de fixation a été proposée. Elle consiste à favoriser tout d'abord la région associée à la tâche principale, puis ensuite une zone adjacente à cette région, située du côté opposé à celui de l'œil qui la regarde pour chaque canal stéréoscopique. Cette stratégie a été incorporée aux méthodes de codage développées dans ce travail.

L'approche de codage proposée consiste à décomposer l'information stéréoscopique en composantes compactes qui sont par la suite combinées afin de former un signal vidéo composite dont l'occupation spectrale n'excède pas celle d'un canal vidéo monoculaire. Les champs gauche et droit de chaque paire stéréoscopique sont décomposés, à l'aide de filtres numériques, en des composantes de luminance passe-bas et passe-haut. Les composantes de hautes-fréquences sont par la suite limitées à une région de fixation, permettant ainsi une réduction de l'espace spectral qu'elles occupent. La sélection de la région de fixation pour chaque canal est faite en fonction

des résultats de notre étude psychophysique, de façon à maximiser la probabilité que le point de fixation de l'observateur se trouve effectivement dans la région de fixation estimée. A l'aide de manipulations spectrales, les composantes de luminance sont combinées à celles de chrominance afin de former un signal vidéo composite dont l'occupation spectrale n'excède pas celle d'un signal vidéo monoculaire. Les *trous spectraux de Fukinuki* sont exploités afin de positionner les composantes de chrominance des deux canaux stéréoscopiques dans différents quadrants du plan fréquentiel temporel-vertical. L'approche proposée est aussi compatible avec le standard NTSC dans la mesure où elle utilise la même sous-porteuse de chrominance et la même région spectrale pour les composantes de chrominance. Il est donc possible d'utiliser des décodeurs NTSC standard pour séparer les composantes de luminance de celles de chrominance. Un traitement additionnel doit par la suite être accompli afin de récupérer l'information correspondant aux canaux gauche et droit.

Trois méthodes de codage, toutes basées sur les principes qui viennent d'être décrits, ont été proposées afin d'exploiter pleinement la bande passante de différents systèmes de communication : (1) la méthode *studio* pour laquelle la limite spectrale horizontale correspond à la moitié de la fréquence d'échantillonnage des images originales, soit 7.16 MHz, (2) la méthode *diffusion* pour laquelle la limite spectrale horizontale est fixée à 4,2 MHz, et (3) la méthode *intégrée* qui peut satisfaire simultanément les contraintes de bande passante des deux premières méthodes. Le format *studio* est approprié pour une utilisation avec, entres autres, des enregistreuses numériques

D2 pour la vidéo composite, alors que le format *diffusion* s'applique surtout à la transmission d'images.

Le niveau de qualité des images codées par les méthodes proposées a été évalué de façon subjective selon un critère de visibilité d'artefacts par quelques sujets et de façon objective à l'aide d'une mesure du rapport signal à bruit. Les résultats sont satisfaisants et démontrent une meilleure qualité des images dans la région de fixation, tel que désiré.

Les principales contributions de ce travail sont les suivantes : (1) l'exploitation d'une région de fixation pour la compression spectrale de séquences, (2) l'étude sur les processus de fixation du système visuel humain, et (3) l'intégration aux méthodes de codage de certains résultats de cette étude chapitre 4. Cette intégration illustre à la fois la pertinence de notre investigation psychophysique et augmente la crédibilité des méthodes de codage qui sont proposées. De plus, l'approche proposée se distingue par une compatibilité avec le standard NTSC et par son efficacité du fait qu'elle implique des opérations de filtrage simples et qu'elle ne nécessite pas d'appariement entre les canaux stéréoscopiques.

Le travail qui a été accompli suscite certaines questions qui pourraient être abordées dans de futurs travaux de recherche. En ce qui concerne le développement d'une stratégie d'estimation de région de fixation en fonction des mouvements oculaires des observateurs, notre travail ne représente qu'un point de départ pour une investigation psychophysique plus systématique sur l'effet de différents attributs des distracteurs

lors de l'accomplissement d'une tâche donnée. Cette investigation pourrait par la suite être étendue à l'accomplissement de différentes tâches afin d'identifier certains comportements de visionnement communs à ces tâches, qui pourraient par la suite être pris en compte pour l'élaboration de stratégies d'estimation de région de fixation. Finalement, une étude plus appliquée pourrait être accomplie afin de voir jusqu'à quel point les résultats obtenus avec des conditions d'études conçues pour faciliter l'étude de certains mécanismes de la vision humaine (stéréogrammes) s'appliquent à des conditions naturelles où un nombre important d'attributs visuels sont simultanément présents.

En ce qui concerne les méthodes de codage, leur amélioration consisterait tout d'abord à éliminer les artefacts ayant été identifiés lors de l'évaluation subjective de la qualité des images. Certaines améliorations ont d'ailleurs déjà été proposées à l'annexe E. Pour une implantation matérielle, une optimisation des filtres devrait être envisagée pour réduire le nombre de coefficients et les exprimer en fonction de nombres étant des multiples d'une puissance de deux. Finalement, tel que mentionné à la fin du chapitre 4, puisque la représentation par composante qui est proposée dans ce travail n'est pas spécifique à une transmission analogique de l'information, il pourrait être intéressant de l'utiliser pour transmettre cette information sous forme numérique.

Bibliographie

- AACH, T. et KAUP, A. (1992), Estimation of physical disparity fields and their discontinuities for stereoscopic images, dans *Proceedings of the European Workshop on 3D-TV Signal Processing*, Rennes, France.
- ABBOTT, A. (1992), A survey of selective fixation control for machine vision, *IEEE Control Systems* **12**(4), 25–31.
- ACCAME, M. et DE NATALE, F. (1995), An adaptive size block-matching approach for stereo image coding, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce. pp. 70–74.
- AGGARWAL, J., DAVIS, L. et MARTIN, W. (1981), Correspondence processes in dynamic scene analysis, *Proceedings of the IEEE* **69**(5), 562–572.
- AHUJA, N. et ABOTT, A. (1993), Active stereo: Integrating disparity, vergence, focus, aperture, and calibration for surface estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(10), 1007–1029.

- ALPARONE, L., ARGENTI, F. et CAPELLINI, V. (1993), A robust coarse-to-fine least squares stereo matching, dans *Proceedings of the 4th European Workshop on Three-Dimensional Television*, Rome, Italie, pp. 171-177.
- ALTUNBASAK, Y., TEKALP, A. et BOZDAGI, G. (1994), Simultaneous motion-disparity estimation and segmentation from stereo, dans *Proceedings of the 1st International Conference on Image Processing*, Vol. 3, Austin, TX, pp. 73-77.
- ANANDAN, P. (1984), Computing dense displacement fields with confidence measures in scene containing occlusion, dans *Proceedings of the SPIE: Intelligent Robots and Computer Vision*, Cambridge, MA, pp. 184-194.
- ANSTIS, S. (1986), Motion perception in the frontal plane: Sensory aspects, dans K. Boff, L. Kaufman et J. Thomas, éditeurs, *Handbook of Perception and Performance: Sensory Processes and Perception*, Vol. 1, Wiley, New York, pp. 16.1-16.27.
- ARDITI, A. (1986), Binocular vision, dans K. Boff, L. Kaufman et J. Thomas, éditeurs, *Handbook of Perception and Performance: Sensory Processes and Perception*, Vol. 1, Wiley, New York, pp. 23.1-23.40.
- AYACHE, N. (1988), Construction et fusion de représentations visuelles. Applications à la robotique mobile, Thèse de doctorat, Université de Paris-Sud, Centre d'Orsay.

- AYACHE, N. et FAVERJON, B. (1987), Efficient registration of stereo images by matching graph descriptions of edge segments, *International Journal of Computer Vision* pp. 107-131.
- BALLARD, D. (1990), Animate vision, rapport technique Comp. Sci. Tech. Rep. 329, Rochester University.
- BARNARD, S. et THOMPSON, W. (1980), Disparity analysis of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2(4), 333-340.
- BROWSE, R. et RODRIGUES, M. (1987), Propagation of interpretation based on graded resolution input, dans *Proceedings of the 1st International Conference on Computer Vision*, pp. 405-410.
- BURT, P. (1988), Smart sensing within a pyramid vision machine, *Proceedings of the IEEE* 76(8), 1006-1015.
- CALIFANO, A., KJELDSSEN, R. et BOLLE, R. (1990), Data and model driven foveation, dans *Proceedings of the 10th International Conference on Pattern Recognition*, pp. 1-7.
- CHASSAING, F. (1995), Digital compatible transmission of a stereoscopic television signal, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 58-63.

- CHASSAING, F., CHOQUET, B. et PELÉ, D. (1991), A stereoscopic television system (3D-TV) and compatible transmission on a MAC channel (3D-MAC), *Signal Processing: Image Communication* 4, 33–43.
- CHASSAING, F. et PELÉ, D. (1991), 3DTV compatible transmission of stereoscopic television signal, dans *Proceedings of the 1st International Festival of 3D Images*, Paris, France, pp. 223–231.
- CHASSAING, F. et PELÉ, D. (1992), Digital and MAC compatible transmission of a stereoscopic signal, dans *Proceedings of the Workshop on Visual Signal Processing and Communications*, Raleigh, NC, pp. 132–137.
- CHASSAING, F. et PELÉ, D. (1993), Digital and analogical compatible transmission of stereoscopic television signal, dans *Proceedings of the 4th European Workshop on Three-Dimensional Television*, Rome, Italie, pp. 181–189.
- CHEBARO, B., CROUZIL, A., MASSIP-PAIHES, L. et CASTAN, S. (1993), Fusion of the stereoscopic and temporal matching results by an algorithm of coherence control and conflicts management, dans *Proceedings of the 5th International Conference on Computer Analysis of Images and Patterns*, Budapest, Hongrie, pp. 486–493.
- CHUPEAU, B. (1993), A multiscale approach to the joint computation of motion and disparity. Application to the synthesis of intermediate views, dans *Proceedings*

of the 4th European Workshop on Three-Dimensional Television, Rome, Italie, pp. 223–230.

CLARK, J. et FERRIER, N. (1988), Modal control of an attentive vision system, dans *Proceedings of the 2nd International Conference on Computer Vision*, pp. 514–523.

COCHRAN, S. et MEDIONI, G. (1992), 3-D surface description from binocular stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(10), 981–994.

COLLEWIJIN, H. et ERKELENS, C. (1990), Binocular eye movements and the perception of depth, dans E. Kowler, éditeur, *Eye Movements and Their Role in Visual and Cognitive Processes*, Vol. 4 de *Reviews of Oculomotor Research*. Elsevier, chapter 4.

CORMACK, R. et FOX, R. (1986), The computation of disparity and depth in stereograms, *Perception and Psychophysics* 38(4), 375–380.

CORNILLEAU-PÉRÈS, V. et DROULEZ, J. (1993), Velocity-based correspondence in stereokinetic images, *Computer Vision, Graphics, and Image Processing: Image Understanding* 58(2), 137–146.

DE VLEESCHAUWER, D. (1991a), Coding of the disparity field and the correction signal in predictive 3DTV coding, dans *Proceedings of the 1st International Festival of 3D Images*, Paris, France, pp. 213–221.

- DE VLEESCHAUWER, D. (1991*b*), Stereoscopic image processing: Image coding, dans *Proceedings of ITE Annual Convention*, Japon, pp. 623–626.
- DE VLEESCHAUWER, D. (1993), An intensity-based, coarse-to-fine approach to reliably measure binocular disparity, *Computer Vision, Graphics, and Image Processing: Image Understanding* 57(2), 204–218.
- DE VLEESCHAUWER, D. et GOUBERT, L. (1992), The influence of camera parameters on the estimation of the parallax field, dans *Proceedings of the European Workshop on 3D-TV Signal Processing*, Rennes, France.
- DINSTEIN, I., GUY, G., RABANY, J., TZELGOV, J. et HENIK, A. (1988), On stereo image coding, dans *Proceedings of the 9th International Conference on Pattern Recognition*, Vol. 1, Rome, Italie, pp. 357–359.
- DINSTEIN, I., GUY, G., RABANY, J., TZELGOV, J. et HENIK, A. (1989), On the compression of stereo images: Preliminary results, *Signal Processing* 17(4), 373–382.
- DINSTEIN, I., KIM, M. et HENIK, A. (1991), Compression of stereo images using subsampling and transform coding, *Optical Engineering* 30(9), 1359–1364.
- DINSTEIN, I., KIM, M., TSELGOV, J. et HENIK, A. (1989), Compression of stereo images and the evaluation of its effects on 3-D perception, dans *Proceedings of the SPIE: Applications of Digital Image Processing XII*, Vol. 1153, San Diego, CA, pp. 522–530.

- DUBOIS, E. et SCHREIBER, W. (1988), Improvements to NTSC by multidimensional filtering, *SMPTE Journal* **97**(6), 446–463.
- DUGELAY, J. (1992), Estimation du mouvement 3D et application à la télévision en relief, Thèse de doctorat, Université de Rennes I, France.
- DUGELAY, J. et PELÉ, D. (1992), Motion and disparity analysis of a stereoscopic sequence: Application to 3DTV coding, dans *Proceedings of the 6th European Signal Processing Conference*, Vol. 3, Bruxelles, Belgique, pp. 1295–1298.
- EASTMAN, R. et WAXMAN, A. (1987), Using disparity functionals for stereo correspondence and surface reconstruction, *Computer Vision, Graphics, and Image Processing* **39**, 73–101.
- ERIKSEN, C. et ST. JAMES, J. (1986), Visual attention within and around the field of focal attention: A zoom lens model, *Perception and Psychophysics* **40**(4), 225–240.
- ERIKSEN, C. et YEH, Y. (1985), Allocation of attention in the visual field, *Journal of Experimental Psychology: Human Perception and Performance* **11**(5), 583–597.
- FALKENHAGEN, L. (1995), 3-D object-based depth estimation from stereoscopic image sequence, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 81–86.

- FAUBERT, J. (1996), Some optical, sensory, and perceptual factors to consider when viewing 3D displays, dans *Proceedings of the 3rd International Display Workshops: 3D Display Technologies and Human Factors*, Kobé, Japon.
- FAUBERT, J. et VON GRÜNAU, M. (1995), The influence of two spatially distinct primers and attribute priming on motion induction, *Vision Research* **35**(22), 3119–3130.
- FUKINUKI, T. et HIRANO, Y. (1984), Extended definition TV fully compatible with existing standards, *IEEE Transactions on Communications* **32**(8), 948–953.
- GIBSON, J. (1989), *Principles of Digital and Analog Communication*, Macmillan, New York.
- GRAMMALIDIS, N., MALASSIOTIS, S., TZOVARAS, D. et STRINTZIS, M. (1993), 3-D motion estimation and compensation for stereoscopic image sequence coding, dans *Proceedings of the 4th European Workshop on Three-Dimensional Television*, Rome, Italie, pp. 191–197.
- GRIMSON, W. (1985), Computational experiments with a feature based stereo algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(1), 17–34.
- GRIMSON, W. (1986), Sensing strategies for disambiguating among multiple objects in known poses, *IEEE Journal on Robotics and Automation* **2**(4), 196–213.

- GRISWOLD, N. et BELL, W. (1988), Disparity coding—an approach for stereo reconstruction, dans *Proceedings of the SPIE: Digital and Optical Shape Representation and Pattern Recognition*, Vol. 938, Orlando, FL, pp. 109–119.
- HENDRIX, C. et BARFIELD, W. (1995), Presence in virtual environments as a function of visual and auditory cues, dans *Proceedings of the 3rd Virtual Reality Annual International Symposium*, pp. 74–82.
- HILLSTROM, A. et YANTIS, S. (1994), Visual motion and attentional capture, *Perception and Psychophysics* 55(4), 399–411.
- HIROSE, M., HIROTA, K. et KIJIMA, R. (1992), Human behavior in virtual environments, dans *Proceedings of the SPIE: Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, pp. 548–559.
- HODGES, L. et DAVIS, E. (1993), Geometric considerations for stereoscopic virtual environments, *Presence* 2(1), 34–43.
- HOFF, W. et AHUJA, N. (1987), Extracting surfaces from stereo images: An integrated approach, dans *Proceedings of the 1st International Conference on Computer Vision*, Londres, Angleterre, pp. 284–294.
- HORST, R. (1992), A digital codec for 3D-TV transmission, dans *Proceedings of the International Workshop on HDTV*, Vol. 2, Kawasaki, Japon.

- IAVECCHIA, H. et FOLK, C. (1994), Shifting attention in visual displays: A time course analysis, *Human Factors* **36**(4), 608–618.
- ISNARDI, M. (1988), Exploring and exploiting subchannels in the NTSC spectrum, *SMPTE Journal* **97**(7), 526–532.
- JAIN, A. (1981), Image data compression: A review, *Proceedings of IEEE* **69**(3), 349–389.
- JAIN, A. (1989), *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ.
- JENKIN, M., JEPSON, A. et TSOTSOS, J. (1991), Techniques for disparity measurement. *Computer Vision, Graphics, and Image Processing: Image Understanding* **53**(1), 14–30.
- JENKIN, M. et KOLERS, P. (1988), Some problems with correspondence, dans W. Martin et J. Aggarwal, éditeurs, *Motion Understanding: Robot and Human Vision*, Robotics: Vision, Manipulation and Sensors, Kluwer Academic, Boston, MA, chapter 8, pp. 269–295.
- JENKIN, M. et TSOTSOS (1986), Applying temporal constraints to the dynamic stereo problem, *Computer Vision, Graphics, and Image Processing* **33**, 16–32.

- JORDAN, J. et BOVIK, A. (1991), Using chromatic information in edge-based stereo correspondence, *Computer Vision, Graphics, and Image Processing: Image Understanding* **54**(1), 98–118.
- KIM, Y. et AGGARWAL, J. (1987), Positioning 3-D objects using stereo images, *IEEE Journal on Robotics and Automation* **3**(4), 361–373.
- KOPERNIK, A. et PELÉ, D. (1992), Improved disparity estimation for the coding of stereoscopic television, dans *Proceedings of the SPIE: Visual Communications and Image Processing*, Vol. 1818 pt.3, Boston, MA, pp. 1155–66.
- KOPERNIK, A. et PELÉ, D. (1993a), Multi-mode object-based 3D sequence analysis, dans *Proceedings of the 4th European Workshop on Three-Dimensional Television*, Rome, Italie, pp. 143–147.
- KOPERNIK, A. et PELÉ, D. (1993b), Object-based 3D image sequence analysis using disparity, motion and structure, dans *Proceedings of the International Workshop on HDTV*, Vol. 1, Ottawa, Canada.
- KOST, B. (1990), Irrelevance reduction of the depth information in stereo images, dans *Proceedings of the SPIE: Stereoscopic Displays and Applications*, Vol. 1256, Santa Clara, CA, pp. 280–288.
- KOWLER, E. (1990), The role of visual and cognitive processes in the control of eye movements, dans E. Kowler, éditeur, *Eye Movements and Their Role in Visual*

and Cognitive Processes, Vol. 4 de *Reviews of Oculomotor Research*, Elsevier, chapter 1.

KROTKOV, E. (1987), Exploratory visual sensing for determining spatial layout with an agile stereo camera head, Thèse de doctorat, Pennsylvania University.

LABERGE, D. et BROWN, V. (1989), Theory of attentional operations in shape identification, *Psychological Review* **96**(1), 101–124.

LAVIE, N. et TSAL, Y. (1994), Perceptual load as a major determinant of the locus of selection of visual attention, *Perception and Psychophysics* **56**(2), 183–197.

LITTLE, J. et GILLETT, E. (1990), Direct evidence for occlusion in stereo and motion, *Image and Vision Computing* **8**(4), 328–340.

LIU, A., THARP, G. et STARK, L. (1992), Depth cue interaction in telepresence and simulated telemanipulation, dans *Proceedings of the SPIE: Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, pp. 541–547.

LIU, J. et SKERJANC, R. (1993), Stereo and motion correspondence in a sequence of stereo images, *Signal Processing: Image Communication* **5**, 305–318.

LLOYD, S., HADDOW, E. et BOYCE, J. (1987), A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling, *Computer Vision, Graphics, and Image Processing* **39**, 202–225.

- MARAPANE, S. et TRIVEDI, M. (1994), Multi-primitive hierarchical stereo analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(3), 227–240.
- MARR, D. (1982), *Vision*, W.H. Freeman, New York.
- MAYHEW, J. et FRISBY, J. (1981), Psychophysical and computational studies towards a theory of human stereopsis, *Artificial Intelligence* **17**, 349–385.
- MORIKAWA, H. et HARASHIMA, H. (1991), 3D structure extraction coding of image sequences, *Journal of Visual Communication and Image Representation* **2**(4), 332–344.
- MUSMANN, H., PIRSH, P. et GRALLERT, H.-J. (1985), Advances in picture coding, *Proceedings of IEEE* **73**(4), 523–548.
- NASRABADI, N. (1992), A stereo vision technique using curve-segments and relaxation matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(5), 566–572.
- NETRAVALI, A. et LIMB, O. (1980), Picture coding: A review, *Proceedings of the IEEE* **68**(3), 366–406.
- NGUYEN, E., LABIT, C. et ODOBEZ, J.-M. (1994), A ROI approach for hybrid image sequence coding, dans *Proceedings of the 1st International Conference on Image Processing*, Austin, TX, pp. 245–249.

- NIKOLAIDIS, N., PITAS, I. et STRINTZIS, M. (1993), Combined evaluation of motion and disparity vector fields for stereoscopic sequence coding, dans *Proceedings of the 5th International Conference on Computer Analysis of Images and Patterns*, Budapest, Hongrie, pp. 387-394.
- OBRADOR, P. (1995), Eye-tracked based stereo video compression scheme, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 64-69.
- OHTA, Y. et KANADE, T. (1985), Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7(2), 139-154.
- OISEL, L., MORIN, L. et LABIT, C. (1995), Stereoscopic data compression using projective geometry, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 46-51.
- OLZAK, L. et THOMAS, J. (1986), Seeing spatial patterns, dans K. Boff, L. Kaufman et J. Thomas, éditeurs, *Handbook of Perception and Performance: Sensory Processes and Perception*, Vol. 1, Wiley, New York, pp. 7.1-7.53.
- OPPENHEIM, A. et SCHAFER, R. (1989), *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, N.J.
- OZKAN, T. et SALARI, E. (1993), Coding of stereoscopic images, dans *Proceedings of the SPIE: Image and Video Processing*, Vol. 1903, San Jose, CA, pp. 228-35.

- PAPADIMITRIOU, D. et DENNIS, T. (1995), Epipolar line estimation and rectification for stereo image pair, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 128–133.
- PATTERSON, R. et MARTIN, W. L. (1992), Human stereopsis, *Human Factors* **34**(6), 669–692.
- PELÉ, D., CHASSAING, F. et KOPERNIK, A. (1992), Analysis and coding of 3DTV, dans *Proceedings of the European Workshop on 3D-TV Signal Processing*, Rennes, France.
- PERKINS, M. (1992), Data compression of stereopairs, *IEEE Transactions on Communications* **40**(4), 684–696.
- POGGIO, T., TORRE, V. et KOCH, C. (1985), Computational vision and regularization theory, *Nature* **317**, 314–319.
- POSNER, M., SNYDER, C. et DAVIDSON, B. (1980), Attention and the detection of signals, *Journal of Experimental Psychology: General* **109**(2), 160–174.
- PRAZDNY, K. (1985), Detection of binocular disparities, *Biological Cybernetics* **52**, 93–99.
- ROBERT, L. et DERICHE, R. (1995), Dense depth map reconstruction using a multiscale regularization approach which preserves discontinuities, dans *Proceedings*

of the International Workshop on Stereoscopic and Three Dimensional Imaging, Santorini, Grèce, pp. 32-39.

SCHERTZ, A. (1992), Source coding of stereoscopic television pictures, dans *Proceedings of the IEE 4th International Conference on Image Processing and its Applications*, Maastricht, Pays-Bas, pp. 462-464.

SCHERTZ, A., FAYAD, M. et BODAMER, G. (1991), Source coding of stereoscopic television pictures, dans *Proceedings of the 1st International Festival of 3D Images*, Paris, France, pp. 165-171.

SEFERIDIS, V. et GHANBARI, M. (1992), Generalized block matching motion estimation, dans *Proceedings of the SPIE: Visual Communications and Image Processing*, Vol. 1818 pt.1, Boston, MA, pp. 110-119.

SEFERIDIS, V. et PAPADIMITRIOU, D. (1993), Improved disparity estimation in stereoscopic television, *Electronic Letters* **29**(9), 782-783.

SELA, G. (1995), Real-time attention for robotic vision, Master's thesis, McGill University.

SHMUEL, A. et WERMAN, M. (1990), Active vision: 3d from an image sequence, dans *Proceedings of the 10th International Conference on Pattern Recognition*, pp. 48-54.

- SPERLING, G. (1990), Comparison of perception in the moving and stationary eye, dans E. Kowler, éditeur, *Eye Movements and Their Role in Visual and Cognitive Processes*, Vol. 4 de *Reviews of Oculomotor Research*, Elsevier, chapter 7.
- STARK, L., EZUMI, K., NGUYEN, T., PAUL, R., THARP, G. et YAMASHITA, H. (1992), Visual search in virtual environments, dans *Proceedings of the SPIE: Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, pp. 577–589.
- STELMACH, L., TAM, W. et HEARTY, P. (1991), Static and dynamic spatial resolution in image coding: An investigation of eye movements, dans *Proceedings of the SPIE: Human Vision, Visual Processing, and Digital Display II*, Vol. 1453, San Jose, CA, pp. 147–152.
- STEURER, J. (1992), Coding stereoscopic images for the human observer, dans *Proceedings of the European Workshop on 3D-TV Signal Processing*, Rennes, France.
- TADENUMA, M. et YUYAMA, I. (1995), Optimization of matching-point-detection in stereoscopic image, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 163–168.
- TAMTAOUI, A. (1992), Coopération stéréovision-mouvement en vue de la compression de séquences d'images stéréoscopiques. Application à la télévision en relief, Thèse de doctorat, Université de Rennes I, France.

- TAMTAOUI, A. et LABIT, C. (1991a), Coherent disparity and motion compensation in 3DTV image sequence coding schemes, dans *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, Toronto, Canada, pp. 2845–2848.
- TAMTAOUI, A. et LABIT, C. (1991b), Constrained disparity and motion estimators for 3DTV image sequence coding, *Signal Processing: Image Communication* 4(1), 45–54.
- TAMTAOUI, A. et LABIT, C. (1991c), TV3D: joined identification of global motion parameters for stereoscopic sequence coding, dans *Proceedings of the SPIE: Visual Communications and Image Processing*, Vol. 1606, Boston, MA, pp. 720–731.
- TAMTAOUI, A. et LABIT, C. (1992), Constrained motion estimators for 3DTV sequence coding, dans *Proceedings of the 6th European Signal Processing Conference*, Bruxelles, Belgique.
- TETSUTANI, N., KISHINO, F. et NAGASHIMA, Y. (1992), Stereoscopic display method employing eye-position tracking and HDTV LCD-projector, dans *Proceedings of the International Workshop on HDTV*, Vol. 2, Kawasaki, Japon.
- TREISMAN, A. (1985), Preattentive processing in vision, *Computer Vision, Graphics, and Image Processing* 31, 156–177.

- TSENG, B. et ANASTASSIOU, D. (1995), Perceptual adaptive quantization of stereoscopic video coding using MPEG-2's temporal scalability structure, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 52-57.
- TZELGOV, J., HENIK, A. et DINSTEIN, I. (1990), Performance consequences of two types of stereo picture compression, *Human Factors* **32**, 173-182.
- TZOVARAS, D., GRAMMALIDIS, N. et STRINTZIS, M. (1995), Depth map coding for stereo and multiview image sequence transmission, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 75-80.
- TZOVARAS, D., STRINTZIS, M. et SAHINOGLU, H. (1994), Evaluation of multiresolution block matching techniques for motion and disparity estimation, *Signal Processing: Image Communication* **6**, 59-67.
- WALDOWSKI, M. (1991), A new segmentation algorithm for videophone applications based on stereo image pairs, *IEEE Transactions on Communications* **39**(12), 1856-1868.
- WAXMAN, A. et DUNCAN, J. (1986), Binocular image flows: steps toward stereomotion fusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6), 715-729.

- WENG, J., AHUJA, N. et HUANG, T. (1988), Two-view matching, dans *Proceedings of the 2nd International Conference on Computer Vision*, Tampa, FL, pp. 64–73.
- WHAITE, P. et FERRIE, F. (1990), From uncertainty to visual exploration, dans *Proceedings of the 3rd International Conference on Computer Vision*, pp. 690–697.
- WILLIAMS, L. et ANANDAN, P. (1986), A coarse-to-fine control strategy for stereo and motion on a mesh-connected computer, dans *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Miami, FL, pp. 219–226.
- XIE, S. (1990), View planning for mobile robots, dans *Proceedings of the International Conference on Robotics and Automation*, pp. 748–754.
- YAMAGUCHI, H., TATEHIRA, Y., AKIYAMA, K. et KOBAYASHI, Y. (1989), Stereoscopic images disparity for predictive coding, dans *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, Glasgow, Ecosse, pp. 1976–1979.
- YARBUS, A. (1967), *Eye movements and vision*, Plenum Press, New York.
- YATES, R., THACKER, N. et LANE, R. (1993), Stretch-correlation for real-time disparity evaluation of stereo pairs, dans *Proceedings of the 4th European Workshop on Three-Dimensional Television*, Rome, Italie, pp. 149–155.

- YEH, Y.-Y. et SILVERSTEIN, L. D. (1992), Spatial judgements with monoscopic and stereoscopic presentation of perspective displays, *Human Factors* **34**(5), 583–600.
- ZHU, Y., GAO, Y. et GOUTTE, R. (1991), Stereo image compression by the combined use of the DCT and binocular fusion and rivalry, dans *Proceedings of the 1st International Festival of 3D Images*, Paris, France, pp. 203–211.
- ZIEGLER, M. (1992), Disparity estimation using variable blocksize, dans *Proceedings of the European Workshop on 3D-TV Signal Processing*, Rennes, France.
- ZIEGLER, M. et PANIS, S. (1995), An object-based stereoscopic coder, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce, pp. 40–45.
- ZIEGLER, M., PANIS, S. et COSMAS, J. (1993), Object-based stereoscopic image coding, dans *Proceedings of the International Workshop on HDTV*, Ottawa, Canada.
- ZIEGLER, M., TENGLER, W. et TABELING, P. (1991), Influence of camera calibration on the coding of stereo sequences, dans *Proceedings of the 1st International Festival of 3D Images*, Paris, France, pp. 183–190.

Annexe A

Principes de codage stéréoscopique

A.1 Compensation de disparité

Deux points se trouvant chacun dans une des images d'une paire stéréoscopique sont *homologues* s'ils proviennent du même point de la scène. La différence de position entre les points correspondants (homologues) est appelée *disparité*. La disparité est un vecteur $\delta(x, y)$ qui exprime le fait que le point homologue au point (x, y) dans une image se situe à $(x + \delta_x(x, y), y + \delta_y(x, y))$ dans l'autre image. Dans la cas d'une configuration de caméras parallèles, la disparité verticale δ_y est égale à zéro. Pour le codage stéréoscopique, plutôt que de coder indépendamment deux images il est possible de transmettre une des deux images ainsi que les vecteurs de disparité permettant de reconstruire la seconde image à partir de l'image transmise. Cette approche de codage s'appelle la *compensation de disparité* et s'exprime selon l'équation A.1 :

$$I_g(x, y) = I_d(x + \delta_x(x, y), y + \delta_y(x, y)) \quad (\text{A.1})$$

Il est possible qu'un vecteur de disparité pointe sur une position qui ne se trouve pas sur la grille d'échantillonnage de l'image reconstruite. Dans ce cas, les valeurs des positions associées à la grille d'échantillonnage sont interpolées à partir des pixels voisins. La reconstruction d'une image par compensation de disparité peut se faire soit par blocs ou soit par pixels lorsqu'un champ de disparité dense est disponible. Pour des séquences stéréoscopiques, des prédictions par compensation de mouvement et par compensation de disparité peuvent être combinées. Quoique ces deux approches reposent sur le même principe, des différences importantes existent entre les vecteurs de mouvement et de disparité, ce qui peut significativement affecter les stratégies de recherche qui sont utilisées pour l'appariement des images.

Premièrement, pour la compensation de mouvement les vecteurs de déplacement peuvent pointer dans n'importe quelle direction, alors que pour la compensation de disparité, en raison de la contrainte épipolaire, la disparité verticale peut être ignorée en adoptant une configuration de caméras parallèles ou en rectifiant les images (Ayache 1988, Papadimitriou et Dennis 1995). Deuxièmement, pour la compensation de mouvement, seul un faible nombre de blocs possèdent généralement un déplacement non-nul entre deux images consécutives, ce qui n'est pas le cas pour la compensation de disparité où la majorité des blocs possèdent une disparité non-nulle. Troisièmement, le déplacement entre deux images successives est généralement faible lorsque le taux

d'échantillonnage temporel est assez élevé, alors que les valeurs de disparité peuvent être beaucoup plus grandes.

Typiquement, pour le codage d'une paire stéréoscopique, la disparité entre les deux images est tout d'abord évaluée, généralement par une méthode d'appariement de blocs (voir la Section A.2.2). Une configuration de caméras parallèles est souvent utilisée afin de limiter l'espace de recherche des blocs correspondants à la direction horizontale. Si l'intervalle des valeurs de disparité est connu *a priori*, l'espace de recherche peut être encore plus réduit (Perkins 1992). L'image à partir de laquelle les vecteurs de disparité sont calculés est par la suite codée selon une méthode de codage monoculaire pour images fixes (Jain 1981, Netravali et Limb 1980), puis est utilisée conjointement avec les vecteurs de disparité pour la prédiction de l'autre image de la paire. Une étude théorique montre que, dans un contexte d'analyse en terme de taux-distorsion, une paire stéréoscopique peut être codée optimalement sans bruit par compensation de disparité en prédisant une image à partir de l'autre image encodée, mais que cette approche est sous-optimale pour du codage avec distorsion (Perkins 1992).

Certaines approches de codage par compensation de disparité remplacent directement les blocs d'une image par leur bloc correspondant dans l'autre image si leur similarité dépasse un certain seuil (Dinstein et al. 1988, Dinstein, Guy, Rabany, Tzelgov et Henik 1989, Ozkan et Salari 1993). L'utilisation d'un seuil de similarité permet de tenir compte des blocs pour lesquels un estimé fiable de la disparité ne peut être

obtenu, en raison d'occlusions par exemple. Certaines autres méthodes combinent un codage prédictif à la compensation de disparité afin d'obtenir une reconstruction de meilleure qualité. En plus des vecteurs de disparité, un signal de correction représentant l'erreur de la prédiction est encodé et transmis. Le signal de correction pour chaque bloc peut être calculé dans le domaine original (Steurer 1992, De Vleeschauwer 1991a, De Vleeschauwer 1991b, Ziegler, Tengler et Tabeling 1991) ou dans le domaine transformé (Perkins 1992). La compensation de disparité n'est pas toujours appliquée à des blocs image. Par exemple, la disparité peut être évaluée entre les contours verticaux des deux images (Griswold et Bell 1988). Si la disparité est obtenue à partir d'un traitement multirésolution, il est possible de transmettre tout d'abord les vecteurs correspondant aux résolutions les plus grossières. Ces vecteurs sont appelés *vecteurs de référence*. Les autres vecteurs qui concernent les résolutions plus fines sont par la suite transmis différentiellement par rapport aux vecteurs de référence.

A.2 Méthodes d'appariement

Le calcul de la disparité demeure un problème difficile (Jenkin et Kolers 1988) en raison des déformations de perspective, de la présence de bruit, de différences dans les conditions d'éclairage et d'occlusions (Little et Gillett 1990). Il existe deux catégories principales d'approches d'appariement: les approches par région et les approches par caractéristiques.

A.2.1 Appariement de caractéristiques

Les approches par caractéristiques consistent à détecter certaines caractéristiques des images et à les mettre en correspondance. Les ambiguïtés d'appariement peuvent être relativement peu nombreuses à condition que les caractéristiques des images soient en quantité relativement faibles et facilement identifiables par des attributs spécifiques (intensité, orientation, courbure, *etc.*) (Ayache et Faverjon 1987, Jordan et Bovik 1991, Weng, Ahuja et Huang 1988). Des contraintes physiques peuvent aussi être utilisées afin de réduire la complexité de la recherche. Par exemple, dans une scène continue avec des surfaces opaques, l'ordonnancement des points homologues le long de droite épipolaires doit être le même dans les deux images. L'existence de ce genre de contrainte permet l'utilisation, entre autres, de méthodes de programmation dynamique (Barnard et Thompson 1980, Lloyd, Haddow et Boyce 1987, Ohta et Kanade 1985) et de relaxation d'étiquettes (Kim et Aggarwal 1987, Nasrabadi 1992). Les caractéristiques les plus souvent utilisées sont celle associées à la détection de contours. En particulier, les passages par zéro du laplacien d'un filtre gaussien (Marr 1982) sont souvent utilisées par des approches multirésolution qui, en contrôlant la variance du filtre gaussien, accomplissent un traitement itératif allant d'une résolution grossière à une résolution plus fine (Eastman et Waxman 1987, Grimson 1985, Mayhew et Frisby 1981). Les méthodes par caractéristiques produisent des champs de disparité éparses. Des champs de disparité denses peuvent être obtenus par interpolation ou en utilisant une approche d'appariement de régions pour supporter

les résultats de l'appariement de caractéristiques (Cochran et Medioni 1992, Hoff et Ahuja 1987, Marapane et Trivedi 1994).

A.2.2 Appariement de régions

Avec les approches d'appariement de régions, des algorithmes d'appariement de blocs sont généralement utilisés. Dans leur forme la plus simple, ces algorithmes consistent à extraire d'une des deux images un bloc de $N \times N$ pixels et à en faire la corrélation avec les blocs centrés sur la droite épipolaire dans l'autre image (Anandan 1984, Jenkin, Jepson et Tsotsos 1991, Tadenuma et Yuyama 1995). La corrélation est souvent basée sur la distance Euclidienne (Aggarwal, Davis et Martin 1981). Si l'énergie du signal peut être considérée constante dans la région d'analyse, la corrélation peut s'exprimer selon l'équation A.2:

$$R_{II}(x, y, i, j) = \sum_{m, n = -\frac{N}{2}}^{\frac{N}{2}} I_d(x + m, y + n) I_g(i + m, j + n) \quad (\text{A.2})$$

Le meilleur appariement est celui qui maximise cette distance. Des blocs de taille variable et une analyse multirésolution peuvent être utilisés (Accame et DE NATALE 1995, Tzovaras, Strintzis et Sahinoglu 1994, Ziegler 1992). Des transformations linéaires peuvent aussi être appliquées afin de tenir compte des déformations de perspective (Alparone, Argenti et Capellini 1993, Seferidis et Ghanbari 1992, Yates, Thacker et Lane 1993). Lorsque l'hypothèse d'énergie constante n'est pas valide, une normalisation à l'aide de l'équation A.3 est utilisée.

$$R_{II}(x, y, i, j) = \frac{\sum_{m,n=-\frac{N}{2}}^{\frac{N}{2}} I_r(x+m, y+n) I_l(i+m, j+n)}{\sqrt{\sum_{m,n=-\frac{N}{2}}^{\frac{N}{2}} I_l^2(i+m, j+n)}} \quad (\text{A.3})$$

Les fonctions de corrélation sont généralement appliquées aux intensités des images, mais elle peuvent aussi être appliquées à d'autres caractéristiques des images moins sensibles aux conditions d'illumination telles que la dérivée des intensités (Alparone et al. 1993) ou les intensités de différentes bandes de fréquence (Jenkin et al. 1991, Williams et Anandan 1986). L'appariement peut être facilité par le fait que les champs de disparité sont généralement cohérents (Poggio, Torre et Koch 1985, Prazdny 1985), c'est-à-dire qu'en raison de la douceur de la surface des composantes de la scène, des vecteurs de disparité voisins sont généralement similaires. Des fonctions objectif complexes incorporant des mesures de cohérence et de corrélation peuvent donc être définies (Aach et Kaup 1992, De Vleeschauwer 1993, Robert et Deriche 1995). D'autres termes peuvent aussi être introduits dans la fonction objectif afin de compenser les variations d'illumination ou les distorsions optiques (De Vleeschauwer et Goubert 1992).

A.3 Comparaison subjective des approches de codage symétriques et asymétriques

La qualité subjective d'images compressées selon des approches symétriques et asymétriques a été comparée par Tzelgov et al. (1990). La codage symétrique a été

accompli à l'aide de la transformée en cosinus discrète tridimensionnelle et le codage asymétrique avec la méthode proposée par Dinstein et al. (1988) et qui consiste à sous-échantillonner les images d'un des canaux stéréoscopiques. Pour les images codées symétriquement, la réduction d'information est à peu près équivalente dans les deux images, alors que pour celles codées asymétriquement, l'information est hautement réduite dans une des deux images alors qu'elle ne l'est que minimalement dans l'autre. Chaque approche de compression résulte en une réduction équivalente de l'information, mais en une différente répartition de l'information entre les deux images. Le codage avec la transformée en cosinus discrète tridimensionnelle est en accord avec la théorie de fusion (Arditi 1986), alors que le codage avec la méthode de Dinstein et al. (1988) supporte la théorie de la suppression (Arditi 1986).

Deux tâches ont été choisies pour l'évaluation de la qualité des images compressées : la reconnaissance d'objets et la discrimination de profondeur. Deux objets, un cube et un cylindre, étaient présentés aux sujets. Pour la tâche de reconnaissance les sujets devaient indiquer le côté (gauche ou droit) où le cube était présenté et pour la tâche de discrimination de profondeur ils devaient indiquer lequel des deux objets se trouvait le plus près d'eux. Pour fin de comparaison, des images monoculaires et des paires stéréoscopiques non compressées ont aussi été utilisées. Les résultats indiquent qu'il n'y a pas de différence significative entre les différentes conditions pour la tâche de reconnaissance. Pour la tâche de discrimination de profondeur, les réponses pour les conditions avec les images compressées furent plus rapides et plus

précises que pour la condition monoculaire, mais moins précises que pour la condition avec les images non compressées. Pour les images compressées, la performance des sujets pour la condition représentant la théorie de la suppression fut meilleure que celle pour la condition de la théorie de la fusion, principalement pour les essais impliquant la discrimination de petites différences de profondeur.

A.4 Estimation coopérative de la disparité et du mouvement

Soit deux paires d'images stéréoscopique de la même scène à deux instants successifs: les images gauche et droite à l'instant t , $I_g(t)$ et $I_d(t)$, et à l'instant $t + \Delta t$, $I_g(t + \Delta t)$ et $I_d(t + \Delta t)$. Soit $\delta_t(x_d, y_d)$ le vecteur de disparité à l'aide duquel la coordonnée du point (x_g, y_g) peut être obtenue à partir de son point homologue (x_d, y_d) et soit $u_d(x_d, y_d)$ et $u_g(x_g, y_g)$ les vecteurs de mouvement projeté au temps t . La contrainte *stéréocinétique* liant le mouvement et la disparité dans une séquence stéréoscopique peut alors être formulée selon l'équation A.4.

$$\delta_t(x_d, y_d) + u_g(x_g, y_g) = \delta_{t+\Delta t}(x'_d, y'_d) + u_d(x_d, y_d) \quad (\text{A.4})$$

Puisque la disparité et le flux optique sont liés, il est possible de prédire un champ de déplacement à partir d'un autre, ce qui, dans un contexte de compression d'images, peut donner lieu à une réduction importante de la quantité d'information à

transmettre. Les approches d'estimation coopérative du mouvement et de la disparité consistent généralement à calculer un champ de déplacement et à utiliser cette information assortie d'une contrainte de cohérence pour l'estimation de l'autre champ. Si la gamme des valeurs de vitesse ou de profondeur est connue *a priori*, cette information peut être utilisée afin de restreindre l'espace de recherche des vecteurs de déplacement (Jenkin et Tsotsos 1986).

Si les champs de flux optique u_g et u_d sont calculés auparavant, ils peuvent être utilisés pour l'estimation de la disparité. Pour des configurations de caméras stéréoscopiques parallèles, $u_y^g = u_y^d$ pour les vecteurs de mouvement projeté d'une paire de points homologues. Cette contrainte peut être intégrée dans le processus d'estimation de disparité en exigeant que les points homologues satisfassent le plus fidèlement possible cette contrainte (Cornilleau-Pérès et Droulez 1993, Liu et Skerjanc 1993, Waxman et Duncan 1986). Une autre approche consiste à identifier les régions en mouvement à partir des champs de flux optique et à apparier ces régions dans les séquences gauche et droite (Chebaro, Crouzil, Massip-Pailhes et Castan 1993, Tamtaoui 1992, Tamtaoui et Labit 1991b, Tamtaoui et Labit 1991a).

Lorsque la disparité à l'instant t ainsi que les vecteurs de mouvement projeté entre les temps t et $t + \Delta t$ sont connus, la disparité au temps $t + \Delta t$ est directement obtenue à l'aide de l'équation A.4. En supposant un faible mouvement en profondeur (*i.e.* $\frac{dz}{dt} \approx 0$) dans le cas de caméras coplanaires, alors $u_g = u_d$. Cette forte contrainte sur le flux optique permet l'estimation de la disparité au temps $t + \Delta t$ à partir du

flux optique d'une seule image (Chupeau 1993).

Lorsque la disparité est connue au temps t , elle peut être utilisée pour l'estimation conjointe des flux optiques gauche et droit en incorporant une contrainte liant les vecteurs de mouvement de points homologues à une fonction objectif à minimiser (Tamtaoui 1992, Tamtaoui et Labit 1991*b*, Tamtaoui et Labit 1991*a*). Finalement, la contrainte stéréocinétique peut être utilisée pour l'évaluation simultanée de tous les champs de déplacement (Altunbasak, Tekalp et Bozdagi 1994, Nikolaidis, Pitas et Strintzis 1993, Tamtaoui 1992, Tamtaoui et Labit 1991*a*, Tamtaoui et Labit 1991*c*, Tamtaoui et Labit 1992) ou pour vérifier la cohérence globale (spatiale et temporelle) des champs de déplacement qui ont été calculés (Chebaro et al. 1993).

L'obtention de champs de mouvement et de disparité cohérents qui satisfont l'équation A.4 est importante puisqu'elle permet la prédiction d'un champ de déplacement par rapport à un autre tout en minimisant les erreurs de prédiction à transmettre. Dans (Tamtaoui 1992, Tamtaoui et Labit 1991*b*, Tamtaoui et Labit 1991*a*), la contrainte stéréocinétique de l'équation A.4 est directement utilisée pour prédire un champ de déplacement à partir d'une combinaison linéaire des autres champs. Des résultats empiriques montrent que cette approche est appropriée pour l'obtention d'estimés initiaux de vecteurs de déplacement, mais qu'elle ne doit pas être utilisée à l'intérieur d'une boucle de prédiction en raison de l'accumulation des erreurs de prédiction. De même, une approche consistant à apparier des régions définies par les caractéristiques de leur mouvement donne aussi lieu à d'importantes erreurs de

reconstruction, principalement parce que les champs de mouvement gauche et droit sont estimés indépendamment (Tamtaoui 1992, Tamtaoui et Labit 1991*b*, Tamtaoui et Labit 1991*a*). Par contre, des approches consistant à estimer conjointement les champs de mouvement et de disparité donnent souvent lieu à de plus petites erreurs de prédiction et s'avèrent donc plus adéquates pour le codage (Tamtaoui 1992, Tamtaoui et Labit 1991*c*, Tamtaoui et Labit 1992).

Annexe B

Perception stéréoscopique et attention

B.1 Géométrie de la vision stéréoscopique

Pour les études sur le système visuel humain, plutôt que d'utiliser la définition de la disparité qui a été introduite à la section A.1, il est plus pratique de travailler avec la disparité rétinienne qui correspond à la différence d'angle de convergence des yeux entre le point de fixation et le point cible (Arditi 1986, Cormack et Fox 1986). La figure B.1 illustre la géométrie pour le calcul de la disparité rétinienne pour le cas de la convergence asymétrique des yeux. A partir de cette figure, le cas plus simple de la convergence symétrique des yeux peut être obtenu en plaçant les points cible et de fixation (C et F) sur le plan mi-sagittal (défini comme le plan passant par le point

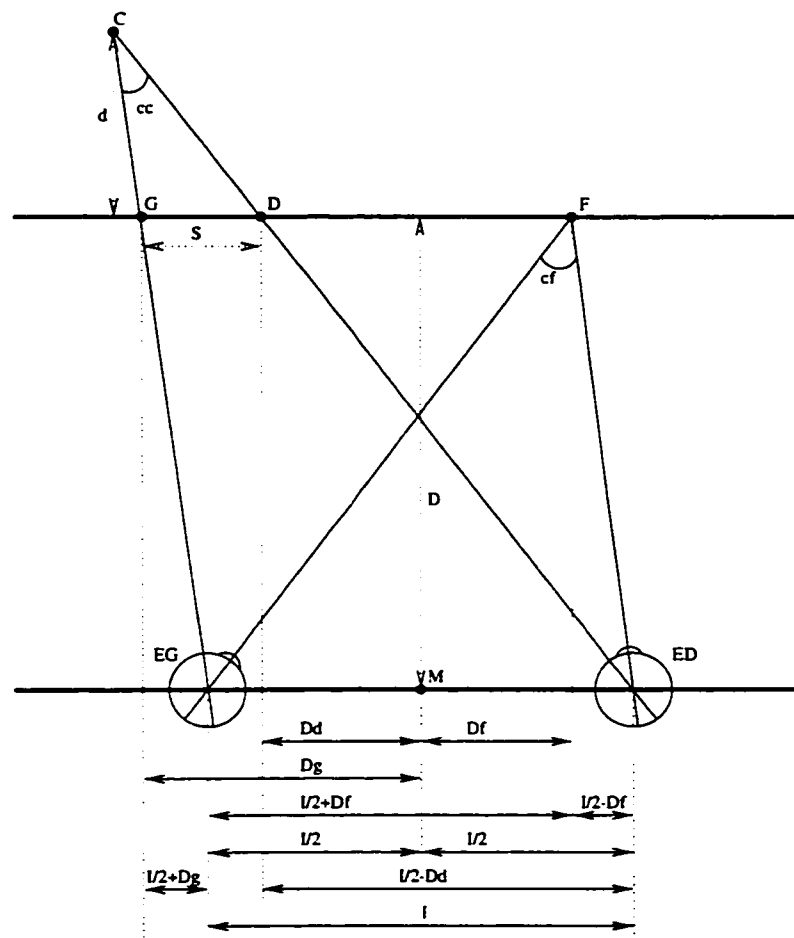


Figure B.1 : Géométrie stéréoscopique pour le calcul de la disparité rétinienne (d'après Cormack et Fox (1986)). EG et ED représentent les yeux gauche et droit, G et D les demi-images gauche et droite, C et F les points cible et de fixation, M le point milieu entre les deux yeux, I la distance interoculaire, S la distance entre les demi-images, cc et cf les angles de convergence au points cible et de fixation, d la distance entre le point cible et le point de fixation, D la distance entre le point de fixation et la ligne joignant le centre optique des yeux et Dg , Dd et Df les distances du plan mi-sagittal aux demi-images gauche et droite et au point de fixation (distances négatives lorsqu'à gauche de M).

milieu M des yeux et perpendiculaire à la ligne joignant les centres optiques des deux yeux). D'après la figure B.1, cf , l'angle de convergence au point de fixation F , est représenté par l'équation:

$$cf = \arctan[(I/2 + Df)/D] + \arctan[(I/2 - Df)/D], \quad (B.1)$$

et cc , l'angle de convergence au point cible C par l'équation :

$$cc = \arctan[(I/2 + Dg)/D] + \arctan[(I/2 + Dd)/D]. \quad (B.2)$$

La disparité rétinienne est la différence entre les deux angles de convergence :

$$r = cf - cc. \quad (B.3)$$

Une disparité rétinienne négative est appelée *disparité croisée* et se produit avec des cibles situées devant le plan de fixation. Une disparité positive, appelée *disparité non-croisée*, se produit avec des cibles situées derrière le plan de fixation. Le calcul de la disparité pour de petits angles de convergence peut être simplifié en supposant que $\arctan(\theta) \approx \theta$. Les équations B.1, B.2 et B.3 deviennent respectivement :

$$cf \approx I/D, \quad (B.4)$$

$$cc \approx I/D + [Dg - Dr]/D, \quad (\text{B.5})$$

$$r \approx (Dg - Dd)/D = S/D. \quad (\text{B.6})$$

De plus amples détails sur la géométrie stéréoscopique peuvent être trouvés dans (Cormack et Fox 1986, Hodges et Davis 1993).

Pour un angle de convergence des yeux, le locus des points dans l'espace ayant une disparité rétinienne nulle est appelé l'*horoptère*. Pour cet angle de convergence, les images gauche et droite d'un objet se situant sur l'horoptère stimulent des points sur la rétine situés aux mêmes distances horizontale et verticale du centre de chaque fovéa. L'horoptère permet donc d'établir une correspondance entre les distances rétiniennes et les positions dans l'espace avec une disparité nulle. En fait, l'horoptère peut être considéré comme le zéro de référence servant à la comparaison de disparités non-nulles pour établir des mesures de profondeur relative. L'horoptère dans le plan horizontal passant par les yeux et qui contient la fovéa est un cercle appelé le cercle de *Vieth-Muller*. Ce cercle est obtenu en supposant que les yeux sont des sphères et que la rotation des yeux se fait autour de l'axe passant par leur nœud optique. Des études empiriques sur la détermination de l'horoptère démontrent que le cercle de Vieth-Muller est une approximation imprécise de la réalité. Une revue des études empiriques sur la détermination de l'horoptère peut être trouvée dans (Arditi 1986).

B.2 Mécanismes de la vision stéréoscopique

Les objets légèrement devant ou derrière l'horoptère produisent de petites disparités. Ces objets sont situés dans une région entourant l'horoptère qui est appelée *l'aire de fusion de Panum*. A l'intérieur de cette région la fusion stéréoscopique est possible et les deux demi-images d'un objet donnent lieu à un percept tridimensionnel (stéréopsie). Les objets situés plus loin de l'horoptère et à l'extérieur de l'aire de fusion de Panum produisent de grandes disparités et des demi-images qui sont habituellement impossible à fusionner. Ces demi-images peuvent être vues comme diplopiques (doubles), mais il est aussi possible qu'un processus de *suppression binoculaire* intervienne afin de réduire la visibilité d'une des deux demi-images. Généralement la visibilité de l'image dominante alterne dans le temps d'un œil à l'autre. Cette alternance est connue sous le nom de *rivalité binoculaire*. La suppression et la rivalité binoculaires ont lieu lors de situation de visionnement normales, mais passent habituellement inaperçues puisque des objets uniques sont généralement perçus.

La stéréopsie peut être qualifiée d'*obligatoire* ou de *qualitative* selon la grandeur de la disparité. La stéréopsie obligatoire concerne la perception de profondeur à partir de petites disparités et d'images majoritairement fusionnées. La profondeur augmente de façon monotone avec la disparité et la valeur de profondeur ainsi que la direction de profondeur (croisée ou non-croisée) sont disponibles. La région de stéréopsie obligatoire s'étend jusque dans la région des images diplopiques. La stéréopsie qualitative concerne la perception de profondeur à partir de grandes disparités et d'images

Tableau B.1 : Valeurs représentatives de limites de disparité pour la fusion binoculaire, la stéréopsie obligatoire et la stéréopsie qualitative pour deux tailles de stimuli et deux valeurs d'excentricité rétinienne (de Patterson et Martin (1992)).

Petite taille (< 15 arcmin)			
	fusion	obligatoire	qualitative
région fovéale	10 arcmin	20 arcmin	2 deg
exc. 6 deg	20 arcmin	2 deg	3,5 deg

Grande taille (1,0–6,6 deg)			
	fusion	obligatoire	qualitative
région fovéale	20 arcmin	2 deg	8 deg
exc. 6 deg	–	–	–

diplopiques et seule la direction de profondeur est disponible. La profondeur perçue n'exhibe donc pas une relation de correspondance monotone avec la disparité.

La plus grande disparité à laquelle la fusion est possible s'appelle la *disparité limite de fusion* et correspond au *seuil de diplopie*, le seuil à partir duquel la fusion est perdue et des images doubles commencent à apparaître. Le *seuil d'acuité stéréoscopique*, quant à lui, indique la plus petite valeur de disparité pour laquelle une discrimination de profondeur est possible. Cette valeur peut être aussi petite que 1,6 arcsec selon la durée de présentation des stimuli (Faubert 1996).

A titre indicatif, le tableau B.1 présente des valeurs représentatives de limites de disparité pour la fusion binoculaire, la stéréopsie obligatoire et la stéréopsie qualitative pour deux tailles de stimuli et deux valeurs d'excentricité rétinienne. Les valeurs du tableau correspondent à la distance entre le stimulus et l'horoptère et s'appliquent autant à des disparités croisées que non-croisées.

B.3 Système attentionnel

Deux paradigmes ont été proposés afin de modéliser les processus d'allocation d'attention : la sélection *hâtive* et la sélection *tardive*. Le paradigme de la sélection hâtive postule que la perception est un processus limité requérant une attention sélective. La sélection attentionnelle se déroule donc rapidement suite à une analyse rudimentaire des caractéristiques d'une image afin de distinguer les stimuli d'intérêt pour la tâche donnée. Conséquemment, les stimuli auxquels l'attention n'est pas portée ne sont pas complètement perçus. Les modèles les plus importants s'inscrivant dans le cadre de ce paradigme utilisent la métaphore d'un faisceau lumineux éclairant une certaine partie du champ visuel — pas nécessairement la fovéa — pour décrire le système attentionnel (Posner, Snyder et Davidson 1980). Une variante de ce modèle utilise plutôt la métaphore d'une loupe grossissante pour laquelle l'intensité du faisceau lumineux varie en fonction de la taille de la région éclairée. Ainsi une petite partie du champ visuel sera éclairée très intensément alors qu'une plus grande région le sera moins (Eriksen et St. James 1986, Eriksen et Yeh 1985). Le modèle du gradient traite surtout des variations de la position de la région d'intérêt qu'il décrit non pas en terme du déplacement d'un faisceau lumineux, mais plutôt en fonction d'une variation continue de la distribution des ressources de traitement (LaBerge et Brown 1989).

A l'opposé, le paradigme de la sélection tardive postule que la perception est un processus illimité qui peut être automatiquement accompli en parallèle dans tout le

champ visuel sans avoir recours à un processus de sélection. La sélection se déroule seulement plus tard, après une perception complète des stimuli, afin de fournir la réponse désirée à une tâche donnée. La théorie de la vision pré-attentive de Treisman s'inscrit dans le cadre de ce paradigme (Treisman 1985). L'observation que certains attributs visuels simples semblent ressortir de l'image, peu importe le nombre de distracteurs qui les entourent, suggère qu'ils sont détectés en parallèle dans tout le champ visuel. Dans la hiérarchie des processus visuels, cette étape de détection se déroulerait avant même que l'attention soit impliquée. Les principaux attributs visuels pouvant être détectés pré-attentivement sont la brillance, la couleur, les terminaisons de lignes, la fermeture, l'inclinaison et la courbure, alors que la détection d'autres attributs tels les intersections, les jonctions, le nombre d'éléments et la connectivité ne se fait pas pré-attentivement. De plus, plusieurs propriétés semblent être codées par rapport à un ou plusieurs points neutres; ce sont seulement les déviations par rapport à ces points qui sont signalées. Par exemple, la courbure et l'inclinaison sont représentées, mais non pas la linéarité et la verticalité.

Récemment, une contrainte additionnelle sur l'allocation des ressources attentionnelles a été proposée afin d'incorporer les deux paradigmes en une théorie unificatrice (Lavie et Tsal 1994). Selon cette théorie, l'attention est toujours une ressource limitée, mais qui doit être allouée complètement. Il est donc impossible d'allouer moins que la capacité totale d'attention disponible à un moment donné. Le facteur déterminant de cette théorie est la charge perceptive associée à une tâche, ce qui correspond à la quan-

tité d'attention requise pour accomplir cette tâche. Ainsi, pour une tâche requérant une grande quantité d'attention, l'attention sélective orientera les ressources d'attention vers les stimuli pertinents à l'accomplissement de la tâche. Par contre, pour une tâche requérant une quantité d'attention moins importante, seulement une partie des ressources attentionnelles sera allouée à la tâche. La partie restante des ressources traitera simultanément et automatiquement l'information non-pertinente à la tâche, c'est-à-dire les distracteurs. L'information non-pertinente sera donc ignorée seulement si l'information prioritaire pertinente à l'accomplissement de la tâche requiert toutes les ressources d'attention disponibles.

B.4 Système oculo-moteur

Les deux plus importants types de mouvement des yeux lors de l'observation d'une scène dynamique sont les mouvements de poursuite douce et les saccades (Kowler 1990). Les mouvements de poursuite douce correspondent à une rotation continue des yeux de façon à suivre un objet en mouvement afin de réduire sa vitesse rétinienne. Ils permettent de diminuer, mais non pas d'abolir le mouvement de l'objet sur la rétine. Leur but est d'amener la vitesse rétinienne de l'objet dans une gamme de vitesses pouvant être traitées par le système visuel humain. Un mouvement de poursuite douce ne peut être initié volontairement en l'absence d'un stimulus en mouvement. La sélection duquel de plusieurs objets en mouvement est poursuivi est déterminé par l'attention sélective. Une fois cet objet choisi, les mouvements de poursuite douce sont

déterminés par rapport aux attentes de l'observateur quant à la trajectoire future de l'objet.

Les saccades sont des mouvements des yeux rapides et brusques permettant de déplacer l'œil d'un point de fixation à un autre. La destination finale d'une saccade peut être contrôlée volontairement par les sujets. L'excentricité rétinienne de la cible affecte la précision spatiale et la latence des saccades. De plus les commandes saccadiques pour atteindre une cible en périphérie peuvent être préparées, complètement ou en partie, avant même l'apparition de la cible. La planification des saccades se fait généralement à l'aide de séquences de mouvement prédéfinies. Les programmes saccadiques peuvent aussi être modifiés rapidement en fonction d'information nouvellement acquise, comme l'apparition d'une nouvelle cible par exemple. L'existence d'un lien entre l'attention sélective et les saccades n'a pas encore été clairement établi. L'attention sélective peut être déplacée sans initier de saccades, mais la question de savoir si des saccades peuvent être accomplies sans déplacer le foyer d'attention demeure ouverte (Kowler 1990).

Il existe aussi d'autres types de mouvements des yeux, comme par exemple les mouvements compensatoires permettant de tenir compte des mouvements de la tête et du corps (Sperling 1990). Le système oculo-moteur contrôle aussi la vergence des yeux de façon à positionner le plan de convergence sur un objet afin que le processus de fusion binoculaire puisse s'accomplir. Contrairement à la croyance populaire, une fois le plan de convergence fixé, les signaux de vergence ne contribuent pas de façon

significative au processus d'inférence de la profondeur (Collewijin et Erkelens 1990).

B.5 Attracteurs d'attention

Il existe plusieurs facteurs visuels qui, dans certaines situations, peuvent attirer l'attention, c'est-à-dire induire un déplacement du foyer d'attention qui peut, à son tour, induire un mouvement des yeux, mais sans que ce soit nécessairement toujours le cas. Souvent, la distinction entre la notion de région d'intérêt et celle de région de fixation n'est pas faite dans la littérature, principalement en raison de la similarité des deux concepts. L'énumération des principaux facteurs pouvant donner lieu à des déplacements d'attention ou de fixations, qui se base principalement sur les revues de la littérature faite par Abbott (1992) et par Sela (1995), présente donc indistinctement les éléments attracteurs d'attention et inducteurs de mouvement des yeux.

- **Attributs visuels de bas-niveau** : en principe, toute caractéristique visuelle détectable peut donner lieu à un déplacement d'attention. En plus des facteurs décrits par Treisman (1985) et rapportés à la section B.3, les régions d'images avec des contrastes élevés ou des fréquences spatiales élevées représentent des attracteurs d'attention (Abbott 1992).
- **Groupement perceptif** : des régions d'images pouvant être définies selon certaines propriétés de groupement telles que le centre de gravité des objets, les sommets de polygones, la symétrie et l'inclusion peuvent induire un changement

d'attention. De plus, lorsque plusieurs objets font partie d'un tout ou qu'un objet est constitué de plusieurs parties, l'attention est dirigée à une position intermédiaire entre ces objets ou ces parties (Sela 1995).

- **Attributs visuels de haut niveau** : lors de l'observation de scènes naturelles, les saccades sont souvent dirigées vers des régions *informatives*, comme par exemple des figures humaines (Abbott 1992).
- **Apparition soudaine** : l'apparition ou la disparition soudaine du stimulus en périphérie peut induire une saccade vers la position du stimulus (Hillstrom et Yantis 1994).
- **Mouvement périphérique** : des objets en mouvement en périphérie représentent des stimuli pouvant donner lieu à des mouvements des yeux vers leur position (Abbott 1992).
- **Proximité de la fovéa** : lorsque plusieurs objets d'intérêt se trouvent dans le champ visuel, ceux se trouvant près de la fovéa représentent des candidats dont les chances d'être choisis pour fixation sont plus élevées. De plus, pour le choix de ces candidats, les mouvements des yeux vers le haut sont favorisés par rapport à ceux vers le bas et un ordre de sélection circulaire par rapport à un point de fixation original est souvent suivi, présumément afin de minimiser la distance totale parcourue lors des déplacements des yeux (Abbott 1992).

- **Présentation stéréoscopique** : pour des situations réelles, le temps de déplacement du foyer d'attention est plus lent pour des objets situés à différents niveaux de profondeur que pour des objets situés au même niveau de profondeur, alors que pour une présentation avec un système stéréoscopique, le temps de déplacement du foyer d'attention ne dépend pas de la profondeur des objets (Iavecchia et Folk 1994).
- **Supplément à une représentation interne** : les mouvements des yeux fourniraient les détails additionnels sur l'environnement qui ne sont pas conservés internement. Puisque le cerveau ne semble pas maintenir une carte spatiale extrêmement détaillée de l'environnement, la fixation des yeux sur différents objets de cet environnement fournirait l'information de haute résolution permettant de suppléer à une représentation interne grossière (Abbott 1992).
- **Tâche** : les séquences de fixation dépendent grandement de la tâche à accomplir. Par exemple, des séquences de fixation très différentes sont observées selon que l'on demande à des sujets d'examiner librement une image, d'estimer l'âge des personnes se trouvant dans cette image ou de trouver la position de certains objets dans l'image (Yarbus 1967).

Annexe C

Données supplémentaires de l'étude sur les stratégies visuelles

C.1 Conditions expérimentales

C.1.1 Installation

Les stimuli ont été présentés sous forme de stéréogrammes dynamiques sur un écran couleur de 14 pouces. Les sujets se trouvaient à 57 cm de l'écran de façon à ce que 1 cm sur l'écran corresponde à un degré d'angle visuel. Une mentonnière a été utilisée afin de stabiliser la position de la tête des sujets. L'éclairage de la pièce était d'environ 40 cd m^{-2} . Les réponses des sujets étaient enregistrées à l'aide d'un clavier d'ordinateur. Les sujets portaient des lunettes avec un filtre rouge sur l'œil gauche et un filtre vert sur l'œil droit. Afin de mesurer le mouvement des yeux, une source

de lumière infrarouge éclairait l'œil gauche et une caméra avec un filtre infrarouge filmait celui-ci. Une deuxième caméra filmait l'écran sur lequel les stimuli étaient présentés. Les deux caméras étaient reliées au système de détection du mouvement des yeux ISCAN (RK-426 et RK-520). Ce système était connecté à un ordinateur qui enregistrerait les points de fixation des sujets.

C.1.2 Stimuli

Les sujets avaient à accomplir une tâche principale et une tâche secondaire. La tâche principale consistait à indiquer en répondant au clavier à l'aide de deux touches numériques si l'objet se trouvant dans la région centrale de l'écran (un cercle ou un carré) approchait ou s'éloignait en profondeur. La Figure C.1 illustre cette tâche. Des distracteurs se déplaçant d'une extrémité à l'autre de l'écran pouvaient interférer avec la tâche principale en masquant les objets de la région centrale durant la partie de leur trajectoire interceptant cette région. La Figure C.2 illustre cette situation. Les distracteurs pouvaient se déplacer selon 8 directions et avec deux types de mouvements : translation dans un plan fronto-parallèle ou translation en s'approchant de l'observateur. Les sujets avaient toutefois la possibilité de faire disparaître temporairement ces distracteurs afin qu'ils n'interfèrent pas avec la tâche principale. La tâche secondaire des sujets consistait donc à détecter la présence de distracteurs et à les faire disparaître au bon moment en appuyant sur la barre d'espacement au clavier afin d'éviter les interférences avec la tâche principale.

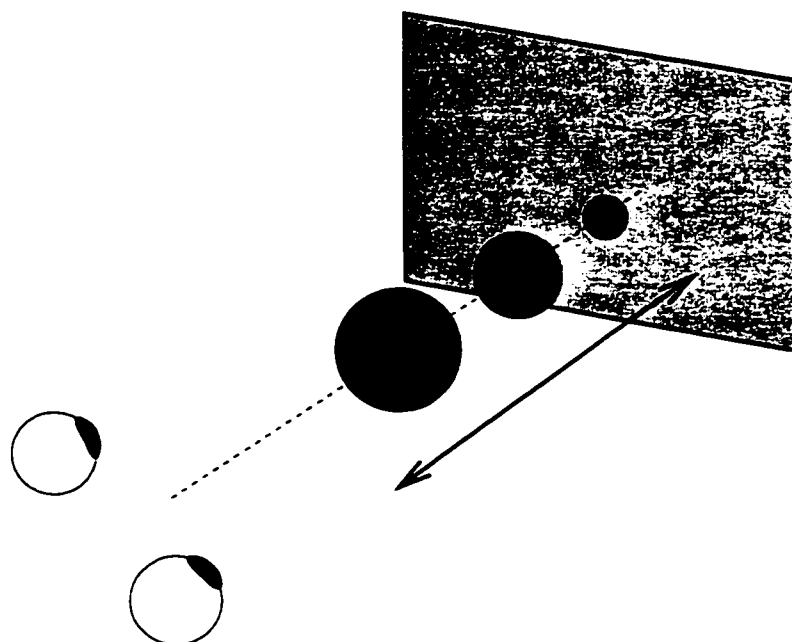


Figure C.1 : Illustration de la tâche principale. Les sujets devaient indiquer si l'objet se trouvant dans la région centrale de l'écran approchait ou s'éloignait en profondeur.

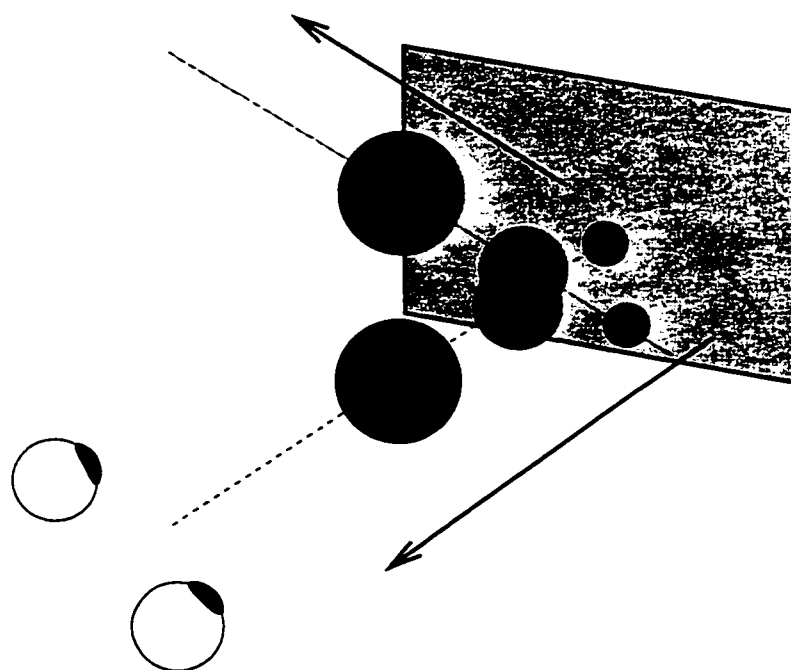


Figure C.2 : Illustration du mouvement des distracteurs. Des distracteurs se déplaçant d'une extrémité de l'écran à l'autre pouvaient interférer avec la tâche principale en masquant les objets de la région centrale durant la partie de leur trajectoire interceptant cette région.

Puisque l'objectif de l'expérience n'était pas de mesurer la performance des sujets lors de l'accomplissement simultanée de deux tâches de détection, les sujets ne pouvaient pas simplement faire disparaître les distracteurs lors de leur apparition dans les régions périphériques de l'image. Une zone d'interaction entre les distracteurs et la région centrale de l'image a donc été définie, et seuls les distracteurs se trouvant à l'intérieur de cette région pouvaient devenir temporairement invisibles. Cette zone d'interaction consistait simplement en une région circulaire englobant la région de la tâche centrale, mais avec un rayon plus grand que celle-ci.

Des points étaient associés à la performance des sujets afin de les motiver à accomplir la tâche. Pour la tâche principale, deux points étaient attribués pour chaque direction de mouvement correcte, un point était retranché pour chaque erreur, et aucun point n'était ajouté ou enlevé en l'absence de réponse. En ce qui concerne la tâche secondaire, un point était retranché si le sujet tentait de faire disparaître le distracteur trop tôt, de façon à l'inciter à ne pas utiliser avec abus la touche de disparition des distracteurs.

C.1.3 Méthode

Douze sujets avec une vision normale ou corrigée à normale ont été testés chacun deux fois. La moitié des sujets ont été testés une première fois pour une condition où les distracteurs avaient des mouvements fronto-parallèles, et une seconde fois où ils avaient des mouvements en profondeur vers l'observateur. L'ordre de présentation de

ces deux conditions a été inversé pour l'autre moitié des sujets. Pour chaque condition, 40 distracteurs ont été présentés, soit 5 pour chacune des 8 directions possibles. La durée d'une présentation était d'environ 3 minutes et demi.

La tâche ainsi que la façon de calculer les points étaient expliquées à chacun des sujets, sans toutefois leur suggérer explicitement de stratégie pour l'accomplissement de la tâche. Tous les sujets étaient d'abord testés afin de vérifier leur capacité à percevoir l'information tridimensionnelle à partir de stéréogrammes dynamiques. Par la suite, le système de mesure du mouvement des yeux était calibré en demandant à chaque sujet de regarder différentes régions à l'écran. Quelques exemples étaient montrés aux sujets afin qu'ils se familiarisent avec la tâche avant que les données soient enregistrées pour fin d'analyse.

C.2 Résultats

C.2.1 Direction du mouvement des distracteurs

Le tableau C.1 présente le pourcentage de fixations dans les zones de la Figure 4.1 pour un sujet lors d'un de ses deux essais, pour les 8 directions des distracteurs exprimées selon les points cardinaux : nord, sud, est, ouest, nord-est, nord-ouest, sud-est et sud-ouest. La moyenne par zone des pourcentages de fixation sur les 8 directions est aussi présentée dans la dernière colonne du tableau. La majorité des fixations se trouvent dans seulement quelques-unes des zones, soit les zones A,B,C et

Tableau C.1 : Pourcentage de fixations d'un sujet dans les zones de la Figure 4.1 pour les 8 directions des distracteurs.

	N	S	E	O	NE	NO	SE	SO	moyenne
A	16,49	2,25	22,14	0	7,35	8,79	3,25	9,14	8,87
B	46,88	44,32	23,56	61,34	42,07	46,44	45,57	57,87	44,90
C	1,17	0	15,84	0	24,62	4,52	0	0	7,29
D	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0,71	0	0	0	0,08
F	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0
H	0,08	0	0	0	0	0,02	0	0	0,02
I	35,39	52,74	37,71	38,66	25,24	39,58	50,45	32,99	38,21
J	0	0	0,14	0	0	0,25	0	0	0,27
K	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0
N	0	0	0,34	0	0	0,05	0	0	0,05
O	0	0	0	0	0	0	0	0	0
P	0	0,68	0,14	0	0	0,16	0	0	0,13
Q	0	0	0	0	0	0,14	0,65	0	0,14
R	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0
T	0	0	0,14	0	0	0,06	0	0	0,04
U	0	0	0	0	0	0	0	0	0

I, peu importe la direction des distracteurs. Ce type de distribution des fixations, concentrées dans seulement quelques-une des zones, est observé pour la plupart des sujets et indique l'absence de corrélation significative entre la direction des distracteurs et la distribution des fixations.

C.2.2 Type de mouvement des distracteurs

Les tableaux C.2, C.3, C.4 et C.5 présentent les pourcentages de fixations dans les zones de la Figure 4.1 pour les deux groupes de sujets, pour toutes les directions de distracteurs combinées. Pour le premier groupe, les distracteurs avec un mouvement dans un plan fronto-parallèle ont été présentés avant ceux avec un mouvement en profondeur vers les observateurs alors que l'ordre de présentation des deux types de distracteurs a été inversé pour le second groupe. Le tableau C.6 présente les différences de pourcentage entre les données moyennes des quatre tableaux. Les deux premières colonnes montrent les différences de pourcentage entre les deux types de distracteurs pour les mêmes sujets et les deux dernières présentent les différences de pourcentage entre les deux groupes de sujets pour le même type de distracteur. Une différence marquée entre l'amplitude des valeurs de différences pour les types de distracteurs différents et celle pour le même type de distracteurs n'est pas observée. En conséquence, il n'est pas possible de conclure à un effet significatif du type de mouvement des distracteurs.

C.2.3 Distributions des fixations

Les données présentées aux tableaux C.2, C.3, C.4 et C.5 ont été classées selon les types de stratégies visuelles utilisées. Afin d'identifier clairement les données auxquelles il est fait référence jusqu'à la fin de cette annexe, celles-ci seront décrites par l'identificateur du sujet suivi entre parenthèse du chiffre 1 ou 2 indiquant le pre-

Tableau C.2 : Pourcentage de fixations dans les zones de la Figure 4.1 pour le premier essai avec les distracteurs possédant un mouvement fronto-parallèle.

	S1	S2	S3	S4	S5	S6	moyenne
A	92,46	90,52	95,50	92,26	63,3	66,23	83,38
B	0,04	1,21	0,39	2,22	0,08	2,86	1,13
C	0	0,66	0,20	0,41	0,10	1,32	0,45
D	0,32	0,52	0,41	0,66	1,23	8,4	1,92
E	0,28	0,01	0,05	0,08	0,11	2,01	0,42
F	0,12	0,04	0,11	0,37	0,23	3,72	0,77
G	0,23	0,43	0,44	0,12	0,72	2,35	0,72
H	5,7	1,37	1,23	0,37	28,74	1,12	6,42
I	0,30	3,41	0,38	2,83	3,44	1,40	1,96
J	0	0,06	0,07	0,04	0,36	1,34	0,31
K	0	0	0,12	0,03	0,07	1,14	0,23
L	0	0,03	0	0,13	0,24	1,44	0,31
M	0	0	0,01	0,08	0,02	1,25	0,23
N	0,02	0,09	0,01	0,03	0	1,73	0,31
O	0	0,03	0,03	0,10	0,10	1,45	0,29
P	0,10	1,05	0,27	0,09	0,25	0,63	0,40
Q	0,25	0,07	0,37	0	0,51	0,88	0,35
R	0	0	0,02	0	0,15	0,22	0,07
S	0	0,03	0,03	0,09	0,08	0,42	0,11
T	0,09	0,43	0,27	0,03	0,12	0,10	0,17
U	0,05	0,02	0,04	0	0,15	0	0,04

Tableau C.3 : Pourcentage de fixations dans les zones de la Figure 4.1 pour le second essai avec les distracteurs possédant un mouvement en profondeur.

	S1	S2	S3	S4	S5	S6	moyenne
A	8,87	96,16	98,48	95,98	74,18	76,31	75,00
B	44,9	0,21	0	0,44	10,17	4,47	10,03
C	7,29	0,45	0	0,78	2,16	2,93	2,27
D	0	1,17	0,19	1,39	0,80	2,49	1,01
E	0,08	0,01	0,01	0,02	0,01	1,42	0,26
F	0	0	0	0,03	0	4,96	0,83
G	0	0,02	0	0,03	0	1,16	0,20
H	0,02	0,71	0,72	0,69	1,55	0,29	0,66
I	38,21	0,42	0,15	0,22	10,29	0,07	8,23
J	0,27	0,07	0	0	0,04	1,61	0,33
K	0	0	0	0	0	0,85	0,14
L	0	0,05	0	0,17	0,20	0,10	0,09
M	0	0	0	0,02	0	0,73	0,13
N	0,05	0,03	0	0,02	0	1,38	0,25
O	0	0,02	0	0	0	0,65	0,11
P	0,13	0,10	0,09	0,09	0,26	0	0,11
Q	0,14	0	0,24	0,07	0,16	0,23	0,14
R	0	0,01	0	0	0	0,23	0,04
S	0	0,07	0	0,04	0,04	0,07	0,04
T	0,04	0,48	0,02	0	0,03	0,06	0,11
U	0	0	0,06	0,02	0,06	0	0,02

Tableau C.4 : Pourcentage de fixations dans les zones de la Figure 4.1 pour le premier essai avec les distracteurs possédant un mouvement en profondeur.

	S7	S8	S9	S10	S11	S12	moyenne
A	91,68	63,97	82,41	82,04	31,38	98,64	75,02
B	0,40	10,42	1,90	5,39	33,12	0,04	8,55
C	0,06	0,96	0,03	0,16	0,80	0,0	0,34
D	0	0,67	1,21	0	0,36	0	0,37
E	0,07	0,03	0,02	0,01	0,01	0,08	0,04
F	0	0,33	0,18	0	0,11	0,41	0,17
G	0,01	0,10	1,27	0	0,02	0	0,23
H	6,34	0,34	1,50	0,64	0,02	0,51	1,56
I	1,18	21,64	7,93	5,71	32,77	0	11,54
J	0	0,48	1,46	1,85	0,72	0	0,75
K	0	0,2	0,24	0,04	0,07	0	0,09
L	0	0,39	0,06	0,03	0	0	0,08
M	0	0,02	0,05	0	0	0,01	0,01
N	0	0	0,04	0	0,21	0	0,04
O	0	0	0,02	0	0,01	0	0,01
P	0,25	0,02	0,28	0	0,06	0,31	0,15
Q	0,02	0,39	1,18	3,98	0,1	0	0,95
R	0	0	0,04	0,04	0,1	0	0,03
S	0	0,02	0,04	0	0,03	0	0,02
T	0	0	0	0	0	0	0
U	0	0	0,07	0,12	0,08	0	0,05

Tableau C.5 : Pourcentage de fixations dans les zones de la Figure 4.1 pour le second essai avec les distracteurs possédant un mouvement fronto-parallèle.

	S7	S8	S9	S10	S11	S12	moyenne
A	96,43	44,22	75,24	99,62	83,10	98,30	82,82
B	1,07	53,78	4,18	0,14	4,84	1,00	10,84
C	0,95	0,81	12,4	0	0	0	2,36
D	0,27	0	3,82	0	0	0	0,68
E	0,10	0	0,35	0,08	0,01	0	0,09
F	0,03	0	0,16	0	0	0,14	0,06
G	0	0,02	0,07	0	0	0	0,02
H	0,19	0,07	0,07	0,14	0,10	0,40	0,16
I	0,04	0,74	2,03	0	10,93	0,11	2,31
J	0,66	0,06	0,23	0	0,06	0	0,17
K	0,09	0,05	0,17	0	0	0	0,05
L	0,04	0,08	0,09	0	0	0	0,04
M	0,02	0,01	0,24	0	0	0,01	0,05
N	0	0	0,03	0	0,01	0	0,01
O	0	0,02	0,05	0	0,03	0	0,02
P	0,08	0,02	0,15	0,02	0,06	0,04	0,06
Q	0	0,07	0,15	0	0,51	0	0,12
R	0,02	0	0,31	0	0,04	0	0,06
S	0	0,03	0,12	0	0,01	0	0,03
T	0	0	0,02	0	0,21	0	0,04
U	0	0,02	0,08	0	0,06	0	0,03

Tableau C.6 : Différence de pourcentage entre les données moyennes des tableaux C.2, C.3, C.4 et C.5

	C.3 - C.2	C.5 - C.4	C.3 - C.4	C.2 - C.5
A	-8,38	7,80	-0,02	0,56
B	8,90	2,29	1,49	-9,7
C	1,82	2,03	1,94	-1,91
D	-0,91	0,31	0,64	1,24
E	-0,16	0,05	0,22	0,33
F	0,07	-0,12	0,66	0,71
G	-0,52	-0,22	-0,03	0,7
H	-5,76	-1,40	-0,90	6,26
I	6,27	-9,23	-3,31	-0,35
J	0,02	-0,58	-0,42	0,14
K	-0,09	-0,04	0,05	0,18
L	-0,22	-0,05	0,01	0,27
M	-0,10	0,03	0,12	0,18
N	-0,06	-0,04	0,21	0,31
O	-0,18	0,01	0,11	0,27
P	-0,29	-0,09	-0,04	0,34
Q	-0,21	-0,82	-0,81	0,23
R	-0,03	0,03	0,01	0
S	-0,07	0,01	0,03	0,08
T	-0,06	0,04	0,11	0,14
U	-0,02	-0,02	-0,03	0,02

mier ou le second essai du sujet. Ainsi par exemple, les données S8(2) réfèrent aux données du sujet S8 lors de son deuxième essai et se trouvent donc à la seconde colonne du tableau C.5. La stratégie consistant à fixer principalement la région centrale de l'écran a été utilisée lors de 12 des 24 essais, soit les essais S1(1), S2(1), S3(1), S4(1), S7(1), S12(1), S2(2), S3(2), S4(2), S7(2), S10(2) et S12(2). Les données représentant la moyenne de ces essais sont présentées à la première colonne du tableau C.7 et illustrées à la Figure 4.2.a. La stratégie consistant à fixer principalement les zones A, B, C et I a été utilisée lors de 9 des 24 essais, soit les essais S8(1), S9(1), S10(1), S11(1), S1(2), S5(2), S8(2), S9(2) et S11(2). Les données représentant la moyenne de ces essais sont présentées à la seconde colonne du tableau C.7 et illustrées à la Figure 4.2.b. Lors d'un des essais d'un des sujets (S5(1)), la somme des pourcentage de fixation dans les zones A et H excède 90% alors que la stratégie d'un autre sujet lors de ses deux essais (S6(1) et S6(2)) consiste à regarder principalement au centre, mais aussi à jeter des coups d'œil dans la plupart des autres régions. Néanmoins, pour ces trois derniers essais, le pourcentage de fixation dans la zone A excède 60%.

C.2.4 Effet d'apprentissage

Une analyse de la performance des sujets ayant utilisé la même stratégie lors de leurs deux essais démontre l'existence d'un effet d'apprentissage se traduisant par une augmentation de la concentration des fixations dans la zone centrale lors du second essai. Cinq sujets ont utilisé la stratégie consistant à fixer principalement la région

Tableau C.7 : Pourcentage moyen de fixations dans les zones de la Figure 4.1 pour les deux principales stratégies visuelles.

	A > 90%	Σ A,B,C,I > 90%
A	95,5	60,6
B	0,60	18,74
C	0,29	2,73
D	0,41	0,76
E	0,07	0,06
F	0,10	0,09
G	0,11	0,16
H	1,5	0,48
I	0,75	14,47
J	0,08	0,57
K	0,02	0,09
L	0,04	0,09
M	0,01	0,04
N	0,02	0,04
O	0,02	0,01
P	0,21	0,11
Q	0,09	0,74
R	0	0,06
S	0,02	0,03
T	0,11	0,03
U	0,02	0,05

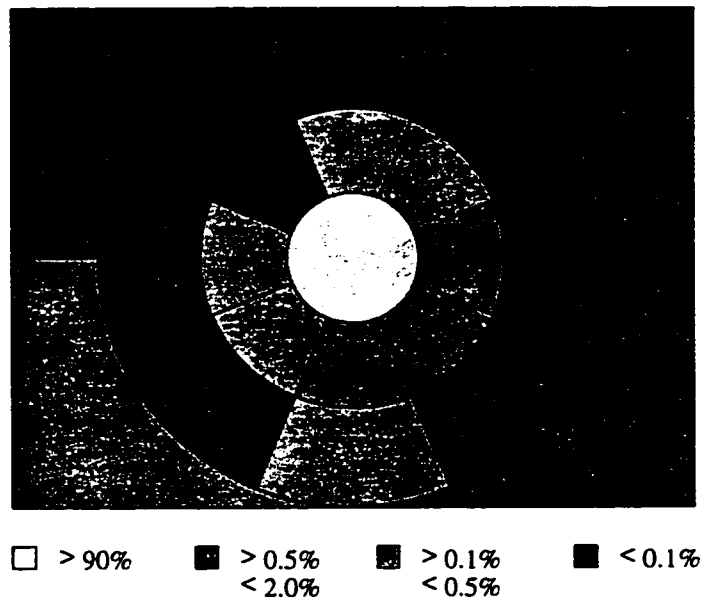


Figure C.3 : Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le premier essai des sujets ayant utilisé la stratégie *centrale* lors de leurs deux essais.

centrale lors de leurs deux essais, soit les sujets S2, S3, S4, S7 et S12. La distribution moyenne de leurs fixations lors de leurs premier et deuxième essais sont illustrées aux Figures C.3 et C.4 et la différence de ces distributions est présentée à la première colonne du tableau C.8 et illustrée à la Figure C.5. Une augmentation moyenne de 3,35% des fixations dans la zone A est observée pour ces cinq sujets. Trois sujets ont utilisé la stratégie consistant à fixer, en plus de la région centrale, les régions se trouvant à sa droite, soit les sujets S8, S9 et S11. La distribution moyenne de leurs fixations lors de leurs premier et deuxième essais sont illustrées aux Figures C.6 et C.7 et la différence de ces distributions est présentée à la seconde colonne du tableau C.8 et illustrée à la Figure C.8. Une augmentation moyenne de 8,27% des fixations dans la zone A est observée pour ces trois sujets.

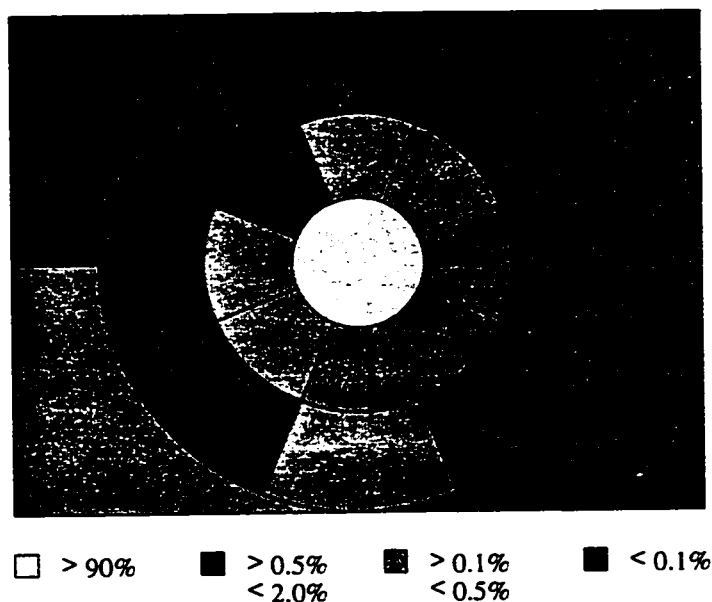


Figure C.4 : Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le second essai des sujets ayant utilisé la stratégie *centrale* lors de leurs deux essais.

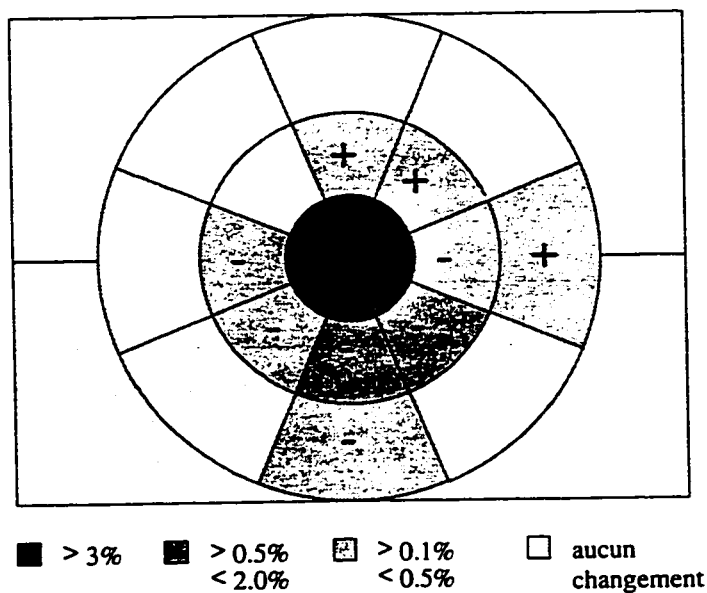


Figure C.5 : Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la stratégie *centrale* lors de leurs deux essais.

Tableau C.8 : Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la même stratégie visuelle lors de leurs deux essais.

	A > 90%	Σ A,B,C,I > 90%
A	3,35	8,27
B	-0,31	5,79
C	0,17	3,81
D	0,29	0,53
E	-0,03	0,10
F	-0,15	-0,15
G	-0,19	-0,43
H	-1,42	-0,54
I	-1,37	-16,2
J	0,11	-0,77
K	-0,01	-0,10
L	0,02	-0,09
M	-0,01	0,06
N	-0,02	-0,07
O	-0,03	0,02
P	-0,31	-0,04
Q	-0,03	-0,31
R	0	0,07
S	-0,01	0,02
T	-0,05	0,08
U	0	0

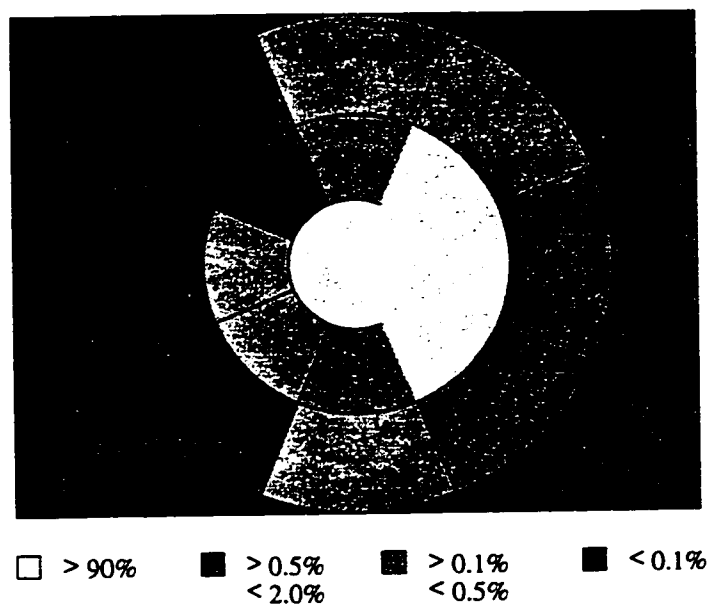


Figure C.6 : Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le premier essai des sujets ayant utilisé la stratégie *centrale + droite* lors de leurs deux essais.

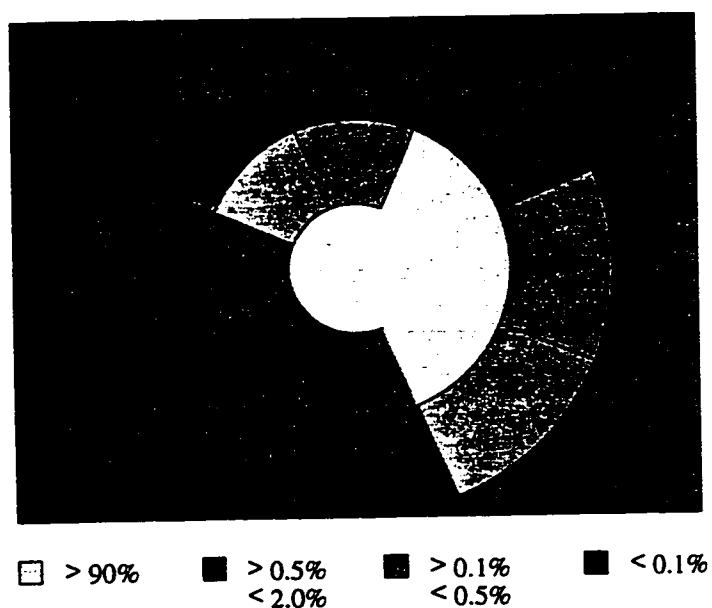


Figure C.7 : Distribution moyenne des fixations dans les zones de la Figure 4.1 pour le second essai des sujets ayant utilisé la stratégie *centrale + droite* lors de leurs deux essais.

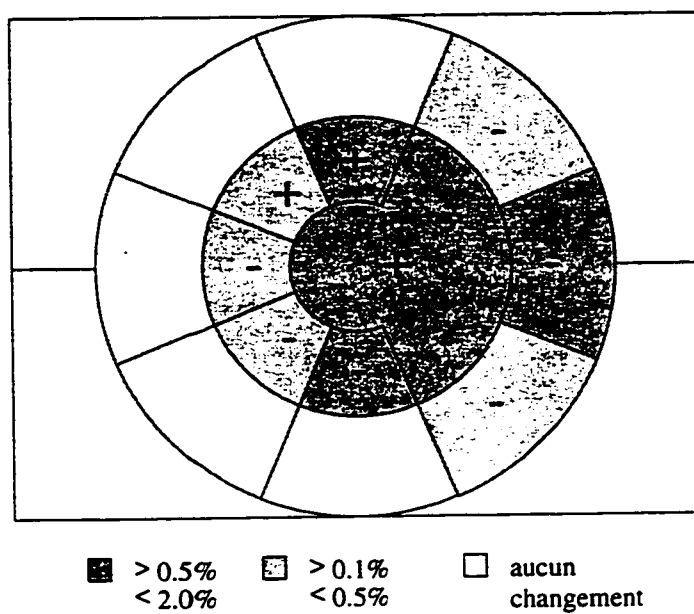


Figure C.8 : Différence de pourcentage de fixations dans les zones de la Figure 4.1 pour les sujets ayant utilisé la stratégie *centrale + droite* lors de leurs deux essais.

Annexe D

Détails supplémentaires sur les méthodes de codage

Cette annexe contient certains détails techniques concernant les méthodes de codage du chapitre 5 qui n'ont pas été insérés dans ce chapitre afin de ne pas alourdir inutilement le texte. Plus spécifiquement, certaines explications sont fournies à la section D.1 quant à la nécessité d'interpoler les composantes de chrominance avant la formation du signal vidéo composite. La conception des filtres spatiaux et temporels qui sont utilisés par les méthodes du chapitre 5 est discutée à la section D.2. La section D.3 décrit la méthode *diffusion* asymétrique. Le traitement pour la reconstruction des composantes avec les méthodes *diffusion* et *intégrée* est présenté aux sections D.4 et D.5. Finalement, le traitement des conditions frontières pour la sélection de la région de fixation est présenté à la section D.6.

D.1 Fréquence de modulation

En raison du format d'échantillonnage spécifié par le standard vidéo NTSC, la taille horizontale des composantes de chrominance est deux fois plus petite que celles des composantes luminance. Par contre, puisque la formation du signal vidéo composite se fait par une addition des composantes de chrominance à celles de luminance, il est nécessaire que celles-ci soient même taille. Cette condition implique donc une interpolation par un facteur de 2 des composantes de chrominance avant la formation du signal vidéo composite et une décimation par le même facteur lors de la reconstruction des composantes transmises. Ces opérations d'interpolation et de décimation sont effectivement accomplies par les méthodes de codage du chapitre 5, excepté lorsque la fréquence de la sous-porteuse est celle du standard NTSC, soit $f_{sp} = 3.58$ MHz. En effet, le choix de cette sous-porteuse permet d'éviter l'interpolation des composantes de chrominance en raison de sa relation avec la fréquence d'échantillonnage des composantes de chrominance.

Supposons que les composantes I et Q aient été échantillonnées à une fréquence de $4f_{sp}$ plutôt qu'à $2f_{sp}$. Le signal de chrominance alors obtenu en modulant en quadrature les composantes I et Q est donné par l'équation D.1 :

$$C(t) = I \cos(2\pi f_{sp}t) + Q \sin(2\pi f_{sp}t), \quad \text{avec } t = \frac{k}{4f_{sp}}, \quad k = 0, 1, 2, \dots \quad (\text{D.1})$$

Selon cette équation, les premières valeurs de C sont $I, Q, -I, -Q, I, Q, -I, -Q, \dots$. Ainsi, seule une valeur sur deux des composantes I et Q est nécessaire à la formation du signal C . Les composantes I et Q échantillonnées à une fréquence deux fois moins élevée, soit $2f_{sp}$, peuvent donc être utilisées sans toutefois nécessiter d'interpolation.

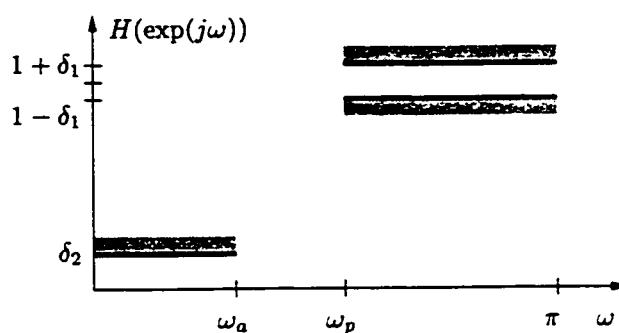
D.2 Filtres

D.2.1 Filtres spatiaux

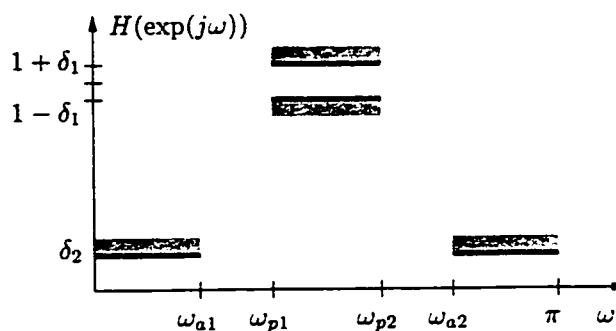
La méthode de Kaiser (Oppenheim et Schafer 1989) a servi pour la conception de tous les filtres spatiaux utilisés par la méthodes du chapitre 5. Les filtres temporels qui servent à l'exploitation des trous de Fukinuki sont décrits à la section D.2.2. La méthode de Kaiser a été choisie parce qu'elle permet de concevoir rapidement des filtres répondant à des spécifications précises avec un faible nombre d'itérations ou de tentatives par essais et erreurs. La conception d'un filtre implique un compromis entre la largeur de la bande de transition, les déviations maximales d'amplitude dans les bandes passantes et d'arrêt et le nombre de coefficients du filtre. Avec la méthode de Kaiser, les déviations maximales d'amplitude dans les bandes passantes et d'arrêt et les fréquences limites de ces bandes doivent être spécifiées. Ces paramètres sont illustrés à la figure D.1 pour des filtres passe-bas, passe-haut et passe-bande. La spécification des fréquences limites des bandes passantes et d'arrêt définit implicitement la largeur de la bande de transition puisqu'elle est égale à $|\omega_p - \omega_a|$, où ω_p et



(a)



(b)



(c)

Figure D.1 : Paramètres pour la spécification d'un filtre : (a) filtre passe-bas, (b) filtre passe-haut et (c) filtre passe-bande (d'après Oppenheim et Schafer (1989)). Les paramètres δ_1 , δ_2 , ω_p et ω_a indiquent respectivement les déviations maximales d'amplitude dans les bandes passante et d'arrêt et les fréquences limites de ces bandes.

ω_a représentent respectivement les fréquences limites des bandes passante et d'arrêt.

Tous les filtres ont été conçus en fixant les paramètres de déviations d'amplitude maximales δ_1 et δ_2 à 0,025 et 0,01 respectivement. Toutefois, la méthode de Kaiser produit des filtres dont l'amplitude de déviation maximale est la même dans les bandes passantes et d'arrêt, soit le minimum des deux paramètres spécifiés, ce qui dans notre cas correspond à une atténuation de 40 dB. Une fois ces paramètres fixés, les fréquences limites des bandes passante et d'arrêt ont été variées de façon à obtenir des filtres avec la largeur de bande de transition, la fréquence de coupure et le nombre de coefficients désirés. Le nombre de coefficients des filtres ont été conçus varie de 11 à 41. Il est important de noter que toutes les opérations de filtrage du chapitre 5 sont accomplies à l'aide de filtres 1D séparables. La figure D.2 illustre un filtre passe-bas de 21 coefficients avec une fréquence de coupure normalisée de 0,5. La figure D.3 illustre l'amplitude de sa réponse fréquentielle et la figure D.4 illustre cette même réponse en décibels. Cette dernière figure illustre bien que l'atténuation du filtre est de 40 dB, tel que spécifié.

Les tableaux D.1, D.2, D.3, D.4, D.5, D.6 et D.7 présentent tous les filtres qui ont été utilisés par les méthodes du chapitre 5. Puisque les filtre à RIF sont symétriques, seule une moitié des coefficients est présentée dans ces tableaux. Les schémas blocs accompagnant la description des méthodes de codage peuvent être consultés afin de déterminer plus précisément l'utilisation de chacun des filtres.

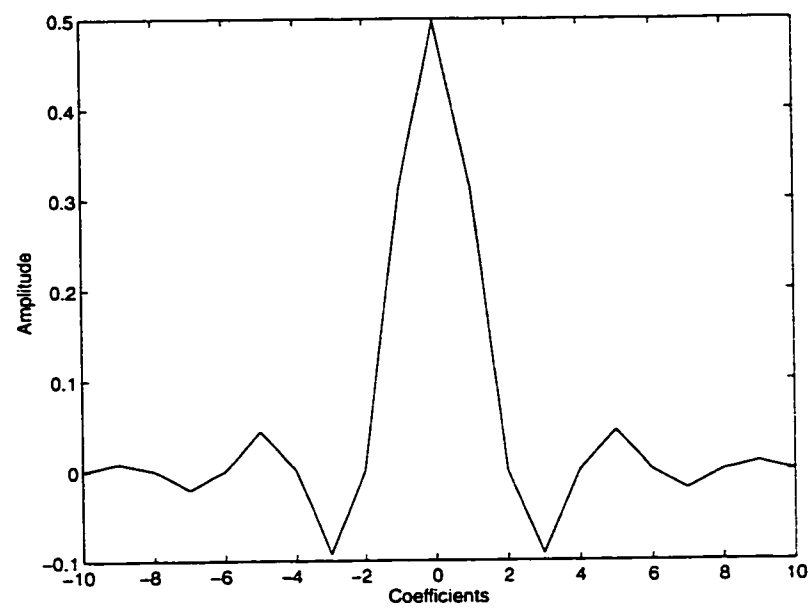


Figure D.2 : Filtre passe-bas de 21 coefficients avec une fréquence de coupure normalisée de 0,5.

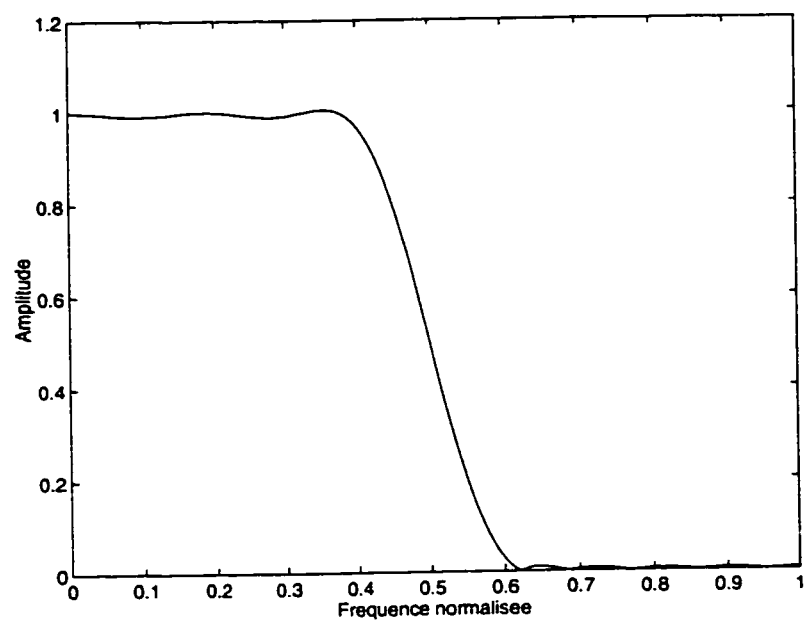


Figure D.3 : Réponse fréquentielle du filtre de la figure D.2.

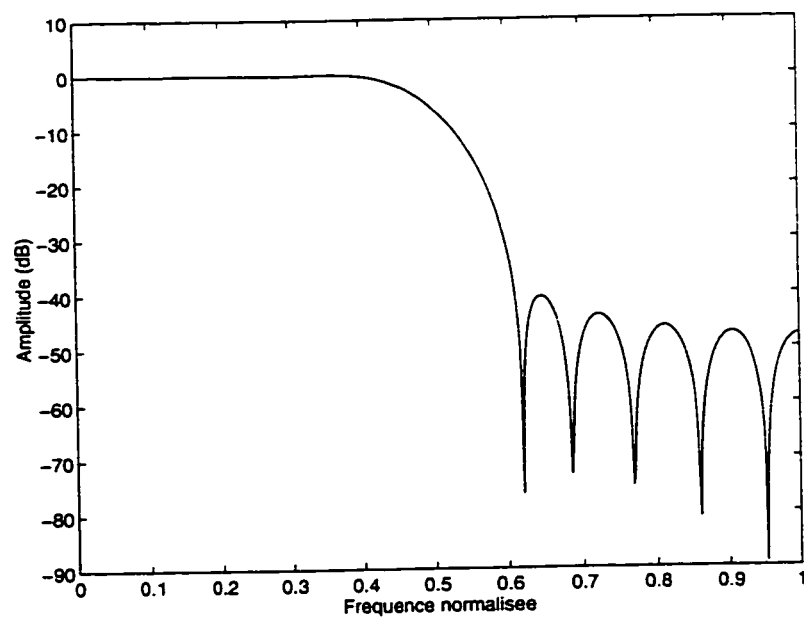


Figure D.4 : Réponse fréquentielle du filtre de la figure D.2 exprimée en décibels.

Tableau D.1 : Filtres passe-bas de luminance pour la subdivision en composantes passe-bas et passe-haut.

	pbas f_{sp}	pbas $f_v/2$	pbas $f_l/2$	pbas $f_v/2$
$h(0)$	0,501144	0,501050	0,291558	0,497935
$h(\pm 1)$	0,317909	0,301252	0,248461	0,312523
$h(\pm 2)$	0,000000	0,000000	0,143922	0,000000
$h(\pm 3)$	-0,102992	-0,061217	0,034123	-0,092789
$h(\pm 4)$	-0,000000	-0,000000	-0,032393	-0,000000
$h(\pm 5)$	0,058331	0,009440	-0,043316	0,043636
$h(\pm 6)$	0,000000		-0,020799	0,000000
$h(\pm 7)$	-0,038143		0,003450	-0,020818
$h(\pm 8)$	-0,000000		0,012130	-0,000000
$h(\pm 9)$	0,026277		0,007663	0,008481
$h(\pm 10)$	0,000000		0,000978	0,000000
$h(\pm 11)$	-0,018355			
$h(\pm 12)$	-0,000000			
$h(\pm 13)$	0,012707			
$h(\pm 14)$	0,000000			
$h(\pm 15)$	-0,008556			
$h(\pm 16)$	-0,000000			
$h(\pm 17)$	0,005484			
$h(\pm 18)$	0,000000			
$h(\pm 19)$	-0,003234			
$h(\pm 20)$	-0,000000			

Tableau D.2 : Filtrés pour le traitement des composantes de luminance passe-haut avec les méthodes *diffusion* et *intégrée*.

	pbas $f_l/3$	pbas $f_l/10$	pbas f_l	pb $f_v/2, 3f_v/4$
$h(0)$	0,201173	0,059145	0,584010	0,172906
$h(\pm 1)$	0,185540	0,058603	0,288285	-0,071031
$h(\pm 2)$	0,143792	0,057000	-0,064886	-0,104055
$h(\pm 3)$	0,089134	0,054407	-0,041662	0,139948
$h(\pm 4)$	0,037218	0,050937	0,024298	-0,019787
$h(\pm 5)$	0,000000	0,046739	0,001960	-0,087323
$h(\pm 6)$	-0,018059	0,041987		0,070490
$h(\pm 7)$	-0,019998	0,036872		0,007173
$h(\pm 8)$	-0,013178	0,031592		-0,039416
$h(\pm 9)$	-0,005035	0,026337		0,016910
$h(\pm 10)$	-0,000000	0,021283		0,004591
$h(\pm 11)$		0,016580		-0,002829
$h(\pm 12)$		0,012351		-0,001289
$h(\pm 13)$		0,008680		-0,003445
$h(\pm 14)$		0,005617		0,005909
$h(\pm 15)$		0,003175		-0,001007
$h(\pm 16)$		0,001336		
$h(\pm 17)$		0,000051		
$h(\pm 18)$		-0,000748		
$h(\pm 19)$		-0,001144		
$h(\pm 20)$		-0,001228		

Tableau D.3 : Filtres passe-bande pour le traitement de la première sous-composante de luminance passe-haut avec les méthodes *diffusion* et *intégrée*.

	pb $0,48f_l, 0,67f_l$	pb $0,67f_l, 0,86f_l$	pb $0,86f_l, 1,05f_l$	pb $1,05f_l, 1,24f_l$
$h(0)$	0,169940	0,170741	0,170570	0,169618
$h(\pm 1)$	0,082493	0,027782	-0,030652	-0,084867
$h(\pm 2)$	-0,080200	-0,149270	-0,147211	-0,075250
$h(\pm 3)$	-0,142402	-0,068607	0,075062	0,142066
$h(\pm 4)$	-0,059293	0,097645	0,091909	-0,066649
$h(\pm 5)$	0,053832	0,076245	-0,081954	-0,045874
$h(\pm 6)$	0,080667	-0,043835	-0,036362	0,080364
$h(\pm 7)$	0,027787	-0,055344	0,057774	-0,034028
$h(\pm 8)$	-0,022307	0,009770	0,004008	-0,017177
$h(\pm 9)$	-0,025853	0,025969	-0,025909	0,025696
$h(\pm 10)$	-0,006311	0,001354	0,003796	-0,008399

Tableau D.4 : Filtres passe-bande pour le traitement de la seconde sous-composante de luminance passe-haut avec les méthodes *diffusion* et *intégrée*.

	pbas $f_l/4$	pb $f_l/4, f_l/2$	pb $5f_l/4, 3f_l/2$	ph $3f_l/2$
$h(0)$	0,169940	0,170741	0,170570	0,169618
$h(\pm 1)$	0,082493	0,027782	-0,030652	-0,084867
$h(\pm 2)$	-0,080200	-0,149270	-0,147211	-0,075250
$h(\pm 3)$	-0,142402	-0,068607	0,075062	0,142066
$h(\pm 4)$	-0,059293	0,097645	0,091909	-0,066649
$h(\pm 5)$	0,053832	0,076245	-0,081954	-0,045874
$h(\pm 6)$	0,080667	-0,043835	-0,036362	0,080364
$h(\pm 7)$	0,027787	-0,055344	0,057774	-0,034028
$h(\pm 8)$	-0,022307	0,009770	0,004008	-0,017177
$h(\pm 9)$	-0,025853	0,025969	-0,025909	0,025696
$h(\pm 10)$	-0,006311	0,001354	0,003796	-0,008399

Tableau D.5 : Filtres passe-bas et passe-haut pour le traitement des composantes de chrominance.

	$\text{pbas } f_v/4$	$\text{ph } 3f_v/4$	$\text{pbas } f_{sp}/7$	$\text{pbas } 3f_{sp}/7$
$h(0)$	0,227155	0,227155	0,138004	0,420637
$h(\pm 1)$	0,205761	-0,205761	0,131712	0,304899
$h(\pm 2)$	0,149882	0,149882	0,114165	0,073525
$h(\pm 3)$	0,080205	-0,080205	0,088960	-0,067580
$h(\pm 4)$	0,019639	0,019639	0,060953	-0,053779
$h(\pm 5)$	-0,016929	0,016929	0,034943	0,012934
$h(\pm 6)$	-0,027476	-0,027476	0,014505	0,030634
$h(\pm 7)$	-0,020521	0,020521	0,001294	0,004385
$h(\pm 8)$	-0,008175	-0,008175	-0,004999	-0,012361
$h(\pm 9)$	0,000675	-0,000675	-0,006120	-0,005634
$h(\pm 10)$	0,003363	0,003363	-0,004416	0,002658

Tableau D.6 : Filtres passe-bande pour le traitement des composantes de chrominance.

	$\text{pb } f_{sp}/2, f_l$	$\text{pb } f_l, 3f_{sp}/2$	$\text{pb } 4f_{sp}/7, 6f_{sp}/7$
$h(0)$	0,322115	0,146314	0,145605
$h(\pm 1)$	0,085589	-0,066781	0,061546
$h(\pm 2)$	-0,222713	-0,078858	-0,086908
$h(\pm 3)$	-0,151028	0,127282	-0,124095
$h(\pm 4)$	0,055720	-0,040178	-0,023098
$h(\pm 5)$	0,063334	-0,067520	0,080338
$h(\pm 6)$	0,001327	0,082065	0,075011
$h(\pm 7)$	0,022390	-0,015528	-0,002706
$h(\pm 8)$	0,025633	-0,040172	-0,048867
$h(\pm 9)$	-0,022635	0,036044	-0,029253
$h(\pm 10)$	-0,028889	-0,002709	0,007444
$h(\pm 11)$	0,000528	-0,013875	0,016698
$h(\pm 12)$	0,004454	0,007922	0,005311
$h(\pm 13)$	-0,001714	0,000092	-0,001997
$h(\pm 14)$	0,004467	-0,000459	-0,000690
$h(\pm 15)$	0,005357	-0,000887	0,000347

Tableau D.7 : Filtres passe-bande pour le traitement des composantes de chrominance et pour la séparation de l'information de chrominance de celle de luminance.

	pb $6f_{sp}/7, 8f_{sp}/7$	pb $f_{sp}/2, 3f_{sp}/2$	pb $f_{sp}/2, 3f_{sp}/2$
$h(0)$	0,145549	0,501320	0,303666
$h(\pm 1)$	0,000000	0,000000	0,244272
$h(\pm 2)$	-0,137403	-0,311177	0,116433
$h(\pm 3)$	0,000000	-0,000000	0,014663
$h(\pm 4)$	0,115172	0,000000	-0,017953
$h(\pm 5)$	-0,000000	-0,000000	-0,009248
$h(\pm 6)$	-0,084653	0,084154	
$h(\pm 7)$	0,000000	0,000000	
$h(\pm 8)$	0,053120	-0,000000	
$h(\pm 9)$	0,000000	0,000000	
$h(\pm 10)$	-0,026855	-0,031813	
$h(\pm 11)$	0,000000	-0,000000	
$h(\pm 12)$	0,009329	0,000000	
$h(\pm 13)$	-0,000000	-0,000000	
$h(\pm 14)$	-0,000693	0,009495	
$h(\pm 15)$	-0,000000	0,000000	

Tableau D.8 : Séquence des composantes de chrominance en fonction de la phase des sous-porteuses principale et auxiliaire. Les variables t , T , ϕ et ϕ' représentent le temps, la période d'un champ, la phase de la sous-porteuse de chrominance et celle de la sous-porteuse auxiliaire.

$t, \phi = 0, \phi' = 0$	$I_g + I_d$	$Q_g + Q_d$	$-I_g - I_d$	$-Q_g - Q_d$
$t + T, \phi = 0, \phi' = 1$	$I_g - I_d$	$Q_g - Q_d$	$-I_g + I_d$	$-Q_g + Q_d$
$t + 2T, \phi = 1, \phi' = 1$	$-I_g - I_d$	$-Q_g - Q_d$	$+I_g + I_d$	$+Q_g + Q_d$
$t + 3T, \phi = 1, \phi' = 0$	$-I_g + I_d$	$-Q_g + Q_d$	$+I_g - I_d$	$+Q_g - Q_d$

D.2.2 Filtres temporels

Afin d'exploiter les trous de Fukinuki, un filtrage temporel doit être accompli au récepteur afin de séparer les composantes de chrominance des deux canaux stéréoscopiques. A l'émetteur, l'amplitude des composantes de chrominance d'un canal est modulée en quadrature par la sous-porteuse f_{sp} et celle des composantes de chrominance de l'autre canal par la sous-porteuse auxiliaire f'_{sp} dont la phase est inversée par rapport à f_{sp} aux deux champs. La relation de phase des deux sous-porteuses est illustrée aux figures D.5 et D.6. A la section D.1, nous avons vu que la modulation d'amplitude en quadrature des composantes de chrominance donnait lieu à une séquence $I, Q, -I, -Q, \dots$. Le tableau D.8 présente la séquence obtenue lorsque les deux sous-porteuses f_{sp} et f'_{sp} sont utilisées pour moduler en quadrature l'amplitude des composantes de chrominance des canaux gauche et droit respectivement.

A partir des valeurs du tableau D.8, il est possible de concevoir des filtres très simples utilisant 3 champs pour séparer les composantes de chrominance du canal

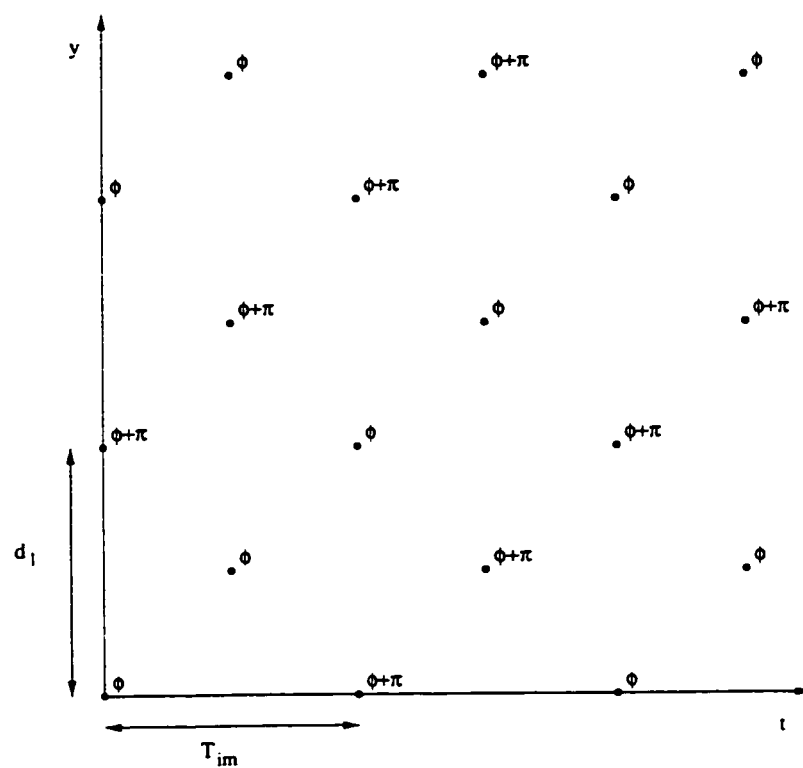


Figure D.5 : Relation de phase de la sous-porteuse f_{sp} . y et t indiquent respectivement les directions verticale et temporelle, T_{ima} représente la période entre une image et d_l la distance entre deux lignes d'un même champ.

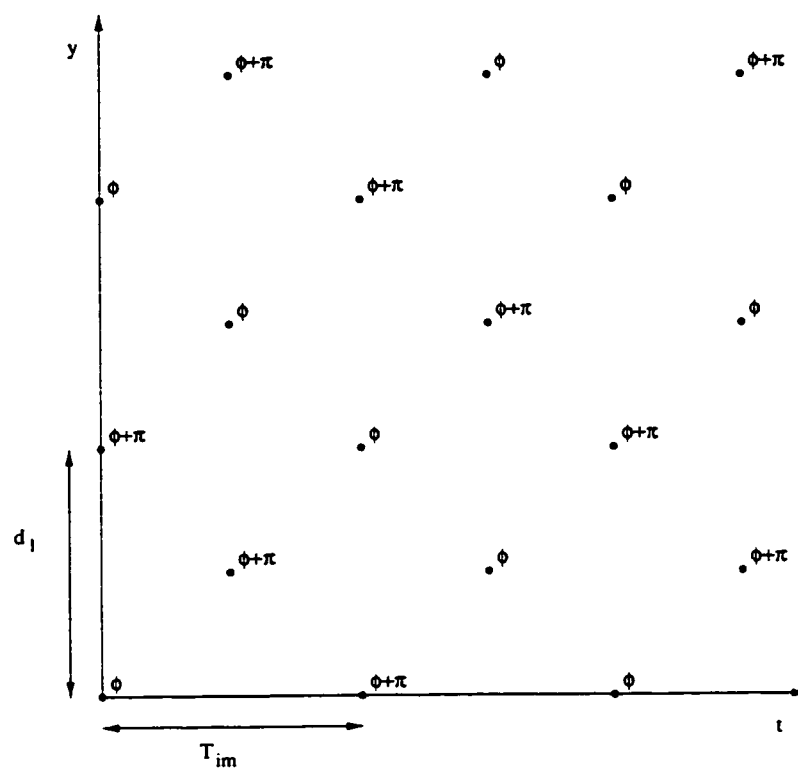


Figure D.6 : Relation de phase de la sous-porteuse f'_{sp} . y et t indiquent respectivement les directions verticale et temporelle, T_{ima} représente la période entre une image et d_l la distance entre deux lignes d'un même champ.

gauche de celles du canal droit. En effet, les filtres $F1 = [-0,5; 0,5; 0]$ et $F2 = [0; 0,5; -0,5]$ appliqués à n'importe quelles trois rangées successives du tableau D.8 (avec $t + 3T$ adjacent à t) permettent d'obtenir les composantes des canaux gauche et droit.

D.3 Méthode diffusion asymétrique

La méthode *diffusion* asymétrique s'inspire de l'observation que la redondance stéréoscopique entre les deux points de vue n'est pas explicitement exploitée par la méthode *studio*. La différence entre les composantes passe-bas des deux canaux stéréoscopiques devrait généralement être petite parce qu'ils correspondent à la même scène observée à partir de deux points de vue légèrement différents. En conséquence, il devrait être avantageux, en terme d'occupation spectrale, de représenter l'information passe-bas des deux canaux par une composante passe-bas et une composante de différence qui occuperont un espace spectral plus compact. Puisque l'objectif principal de la méthode consiste à représenter la région de fixation avec la meilleure qualité possible, les différences entre les deux canaux devraient être minimisées à l'intérieur de cette région. Cette minimisation peut être accomplie en décalant horizontalement (de façon circulaire) le champ d'un canal de façon à aligner la région de fixation.

La méthode asymétrique proposée consiste donc à décomposer l'information de luminance YG et YD des deux canaux en des composantes passe-bas, $YPBG$ et $YPBD$, et en des composantes passe-haut, $YPHG$ et $YPHD$, à l'aide de filtres spa-

tiaux 2D séparables à RIF dont les fréquences de coupure horizontale et verticale sont respectivement 2,2 MHz et $f_v/2$. Après avoir décalé la composante passe-bas droite de façon à aligner les régions de fixation, un signal de différence $\Delta YPBD$ possédant la même largeur de bande que le signal original est formé. Afin de compacter d'avantage l'information transmise, le signal de différence est filtré horizontalement pour limiter sa largeur de bande à 2,0 MHz. La figure D.8 illustre la décomposition de l'information de luminance des deux canaux en une composante passe-bas, une composante passe-bas de différence et deux composantes passe-haut. Les composantes de hautes-fréquences de chaque canal sont ensuite divisées en deux sous-composantes : les hautes fréquences verticales dont les fréquences horizontales sont plus petites que 2,2 MHz, $YPH1G$ et $YPH1D$, et les hautes- fréquences verticales plus élevées que 2,2 MHz, $YPH2G$ et $YPH2D$. Ces composantes sont limitées spatialement à la région de fixation, agrandies horizontalement par un facteur de 5 et verticalement par un facteur de 2, puis filtrées passe-bas afin d'éliminer les répétitions spectrales. Les composantes de luminance ainsi traitées, appelées $YRF1G$, $YRF1D$, $YRF2G$ et $YRF2D$, sont combinées avec les composantes de luminance passe-bas $YPBG$ et $\Delta YPBD$ et celles de chrominance, dont le traitement est le même que pour la méthode *diffusion* symétrique décrite au chapitre 5, par des opérations de modulation et de filtrage de façon à occuper l'espace spectral tel qu'illustré à la figure D.7. Même si cette méthode est adéquate parce que la région de fixation est reproduite de façon satisfaisante, des distorsions perceptibles sont créées dans la région périphérique parce

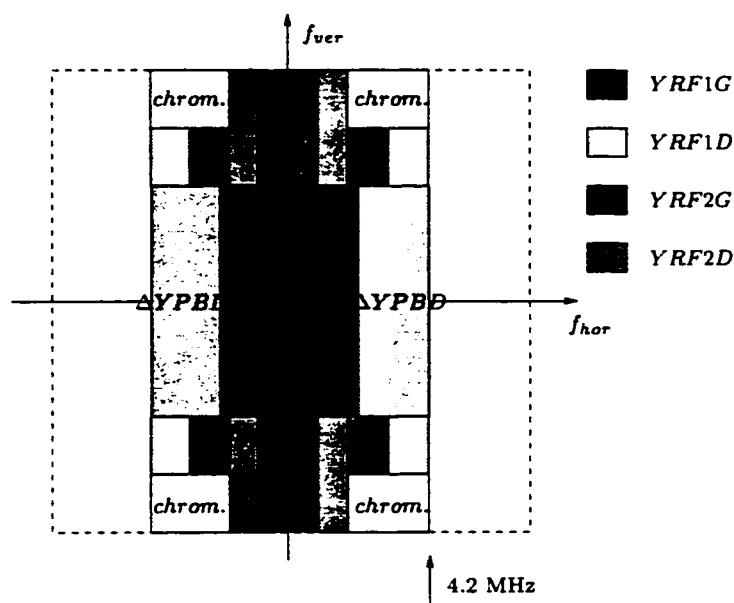


Figure D.7 : Division de l'espace spectral avec la méthode *diffusion* asymétrique. $YPBG$, $\Delta YPBD$, $\{YRF1G, YRF1D, YRF2G, YRF2\}$ et *chrom.* représentent respectivement les composantes de luminance passe-bas gauche, de différence passe-bas droite, les sous-composantes de luminance passe-haut gauche et droite associées à la région de fixation et les composantes de chrominance.

qu'il manque de l'information dans la bande de fréquence allant de 2,0 à 2,2 MHz.

C'est pour cette raison que l'approche *diffusion* symétrique présentée au chapitre 5 est plutôt favorisée.

D.4 Reconstruction avec la méthode diffusion

Les figures D.9 et D.10 illustrent à l'aide de schémas bloc la reconstruction des composantes transmises à partir du signal vidéo composite lorsque les méthodes *diffusion* et *diffusion* avec trous de Fukinuki sont utilisées. Les figures D.11, D.12, D.13

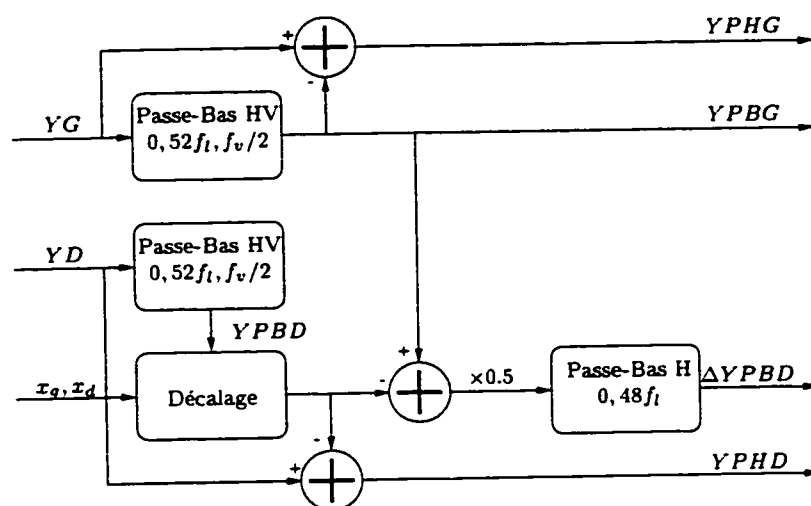


Figure D.8 : Décomposition de l'information de luminance en composantes passe-bas et passe-haut avec la méthode *diffusion* asymétrique. YG et YD représentent les composantes de luminance originales et $YPBG$, $\Delta YPBD$, $YPHG$ et $YPHD$ les composantes de luminance passe-bas gauche, de différence passe-bas droite et passe-haut gauche et droite. x_g et x_d correspondent à la position de la région de fixation dans les canaux gauche et droit.

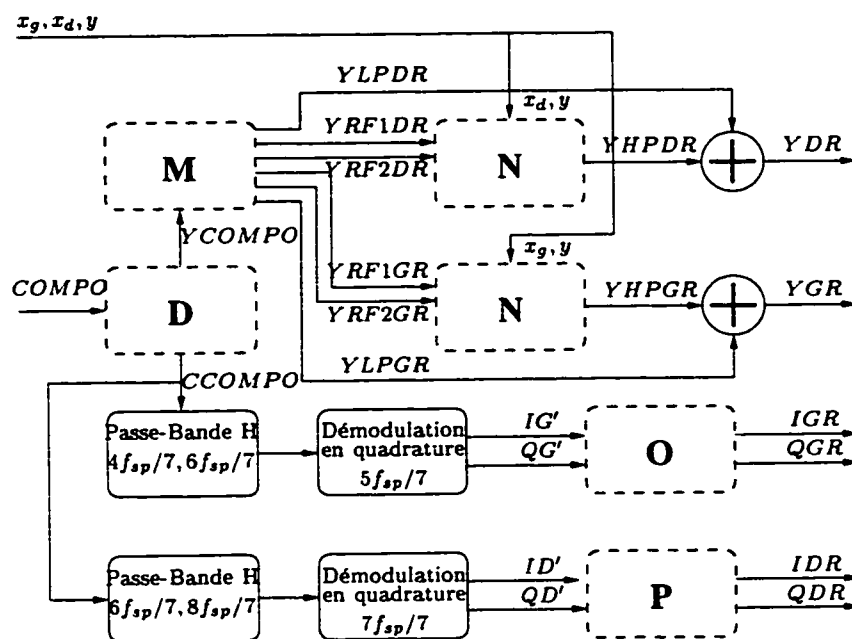


Figure D.9 : Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode *diffusion*. $COMPO$ correspond au signal vidéo composite et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g et x_d représentent la position horizontale de la région de fixation dans les canaux gauche et droit, et y la position verticale de cette région dans les deux canaux. Le bloc **D** se trouve à la figure 5.10, le bloc **M** à la figure D.11, le bloc **N** à la figure D.12, le bloc **O** à la figure D.13 et le bloc **P** à la figure D.14.

et D.14 illustrent de façon plus détaillée les étapes de traitement des figures D.9 et D.10.

D.5 Reconstruction avec la méthode intégrée

Les figures D.15 et D.16 illustrent à l'aide de schémas bloc la reconstruction avec la méthode *intégrée* des composantes transmises à partir du signal vidéo composite lorsque la bande passante du système de transmission d'images est respectivement de

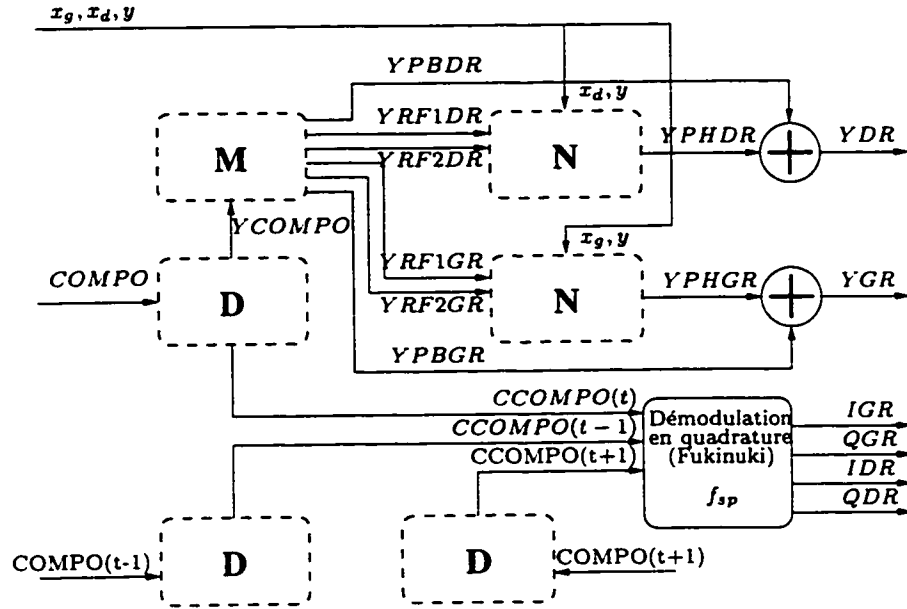


Figure D.10 : Reconstruction des composantes à partir du signal vidéo composite avec la méthode *diffusion* avec trous de Fukinuki. $COMPO(t-1)$, $COMPO(t)$ et $COMPO(t+1)$ représentent respectivement le signal vidéo composite à l'instant précédent, l'instant courant et l'instant suivant. De même $CCOMPO(t-1)$, $CCOMPO(t)$ et $CCOMPO(t+1)$ correspondent à la partie du signal vidéo contenant l'information de chrominance à l'instant précédent, l'instant courant et l'instant suivant. $YCOMPO(t)$ représente la partie du signal vidéo contenant l'information de luminance à l'instant courant et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ correspondent aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g et x_d représentent la position horizontale de la région de fixation dans les canaux gauche et droit, et y la position verticale de cette région dans les deux canaux. Le bloc D se trouve à la figure 5.10, le bloc M à la figure D.11 et le bloc N à la figure D.12.

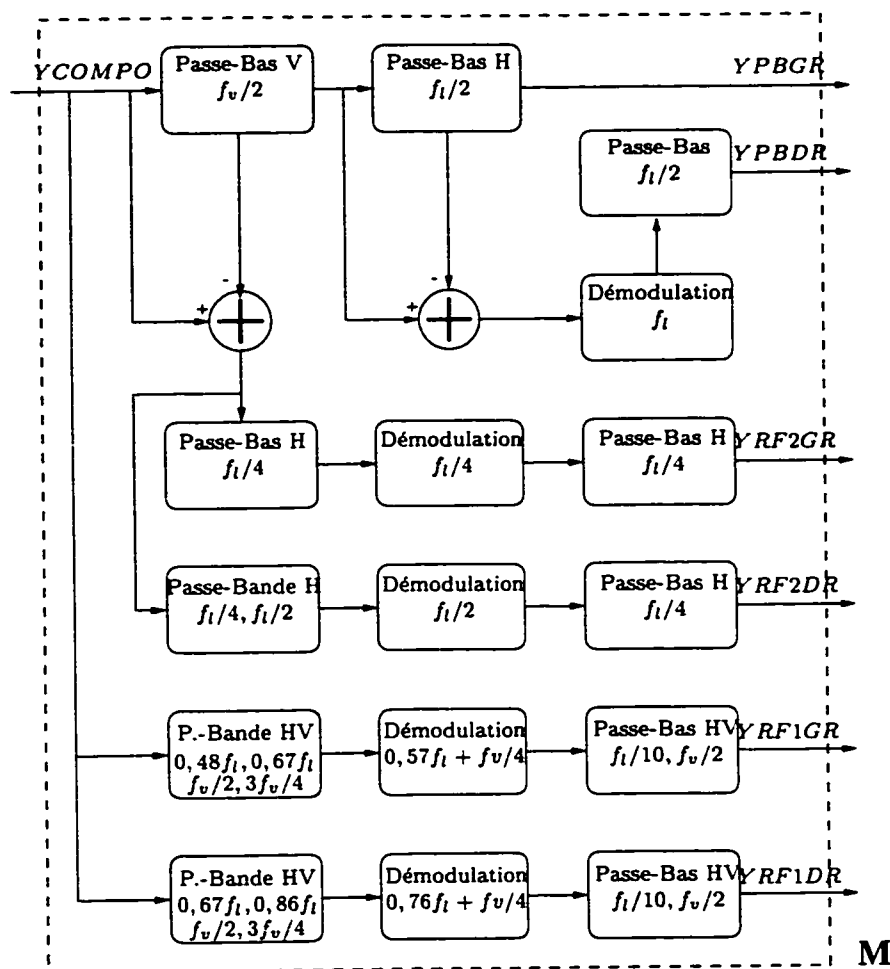


Figure D.11 : Séparation des composantes de luminance avec la méthode *diffusion*. $YCOMPO$ représente la partie du signal vidéo composite contenant l'information de luminance et $\{YLPGR, YPBDR\}$, $\{YRF1GR, YRF1DR, YRF2GR, YRF2DR\}$ les composantes de luminance reconstruites passe-bas et les sous-composantes de luminance reconstruites passe-haut de la région de fixation pour les canaux gauche et droit.

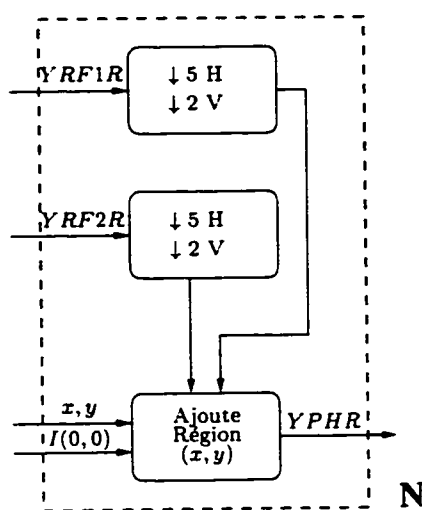


Figure D.12 : Reconstruction des composantes de luminance passe-haut avec la méthode *diffusion*. $YRF1R$, $YRF2R$ et $YPHR$ correspondent respectivement aux deux sous-composantes de luminance reconstruites passe-haut associées à la région de fixation et à la composante de luminance reconstruite passe-haut correspondant à l'image entière. $I(0, 0)$ représente une image de la taille d'un champ dont tous les éléments sont égaux à 0 et x et y correspondent à la position de la région de fixation.

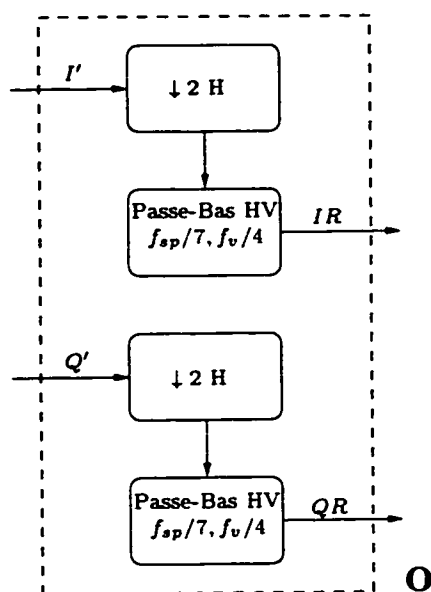


Figure D.13 : Reconstruction des composantes de chrominance du canal gauche avec la méthode *diffusion*. I' , Q' , IR et QR correspondent respectivement aux deux composantes de chrominance démodulées et aux deux composantes de chrominance reconstruites.

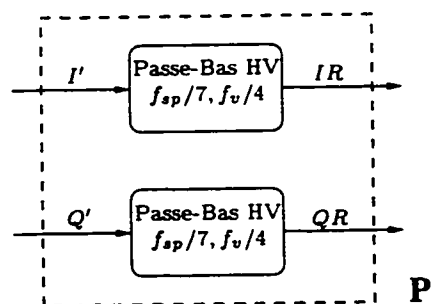


Figure D.14 : Reconstruction des composantes de chrominance du canal droit avec la méthode *diffusion*. I' , Q' , IR et QR correspondent respectivement aux deux composantes de chrominance démodulées et aux deux composantes de chrominance reconstruites.

7,16 MHz et 4,2 MHz. A la figure D.16 le filtrage passe-bas à 4,2 MHz du signal vidéo est illustré afin de faciliter la compréhension de la méthode de reconstruction, même si ce filtrage est normalement accompli à l'émetteur avant la transmission du signal. Les figures D.17 et D.18 illustrent de façon plus détaillée les étapes de traitement des figures D.9 et D.10.

D.6 Conditions frontières pour la sélection de la région de fixation

Puisque seule la position de la région de fixation est spécifiée et que la taille de celle-ci est fixe, mais de grandeur variable selon la méthode de codage utilisée, il est possible que cette région dépasse les limites de l'image. Dans le cas des méthodes *studio* et *diffusion*, lorsqu'un tel débordement se produit, la région de fixation est choisie de façon à ce que sa taille corresponde à la taille maximale allouée à la région de fixation. Par exemple, dans le cas d'un débordement horizontal vers la gauche, les coordonnées horizontales de la région de fixation iront de 1 à $L/2$ ou de 1 à $L/5$ selon la méthode de codage, avec L représentant la largeur totale de l'image. De la même façon pour un débordement horizontal vers la droite, les coordonnées de la région de fixation iront de $L - L/2$ à L ou de $L - L/5$ à L . Le même principe s'applique aussi pour un débordement dans la direction verticale. Il est important d'observer qu'avec la méthode *studio*, un décalage de la région de fixation par rapport

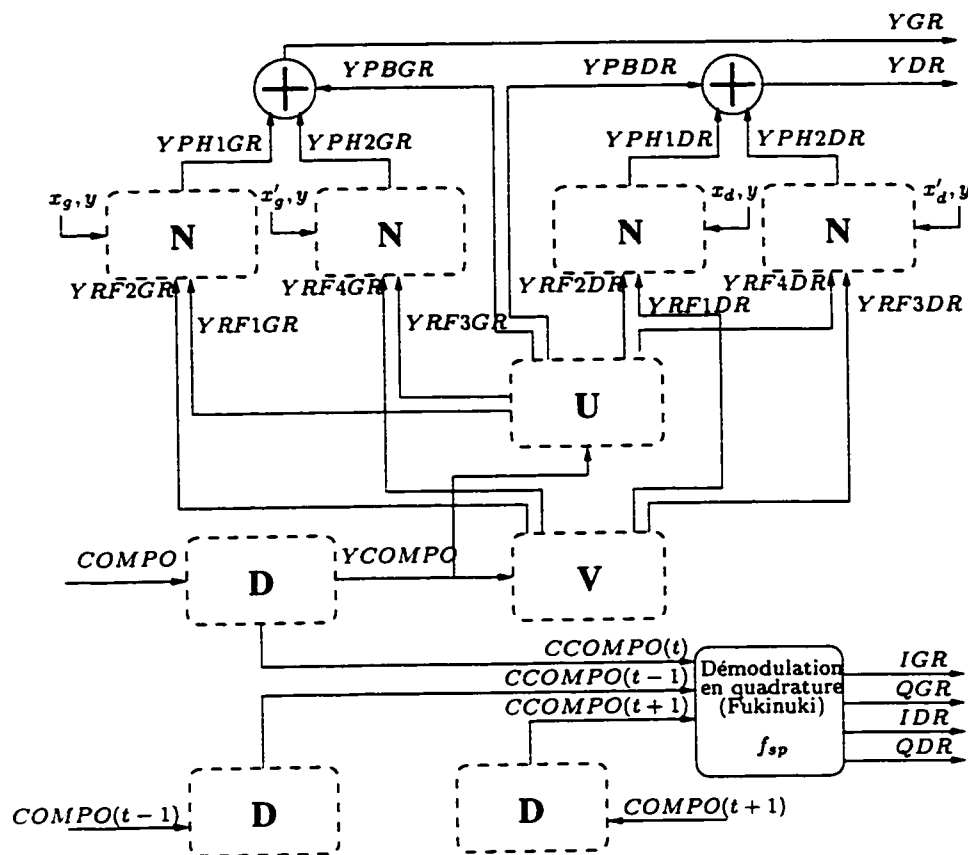


Figure D.15 : Reconstruction des composantes à partir du signal vidéo composite avec la méthode *intégrée* lorsque la largeur de bande du signal est de 7,16 MHz. $COMPO(t-1)$, $COMPO(t)$ et $COMPO(t+1)$ représentent respectivement le signal vidéo composite à l'instant précédent, l'instant courant et l'instant suivant. De même $CCOMPO(t-1)$, $CCOMPO(t)$ et $CCOMPO(t+1)$ correspondent à la partie du signal vidéo contenant l'information de chrominance à l'instant précédent, l'instant courant et l'instant suivant. $YCOMPO(t)$ représente la partie du signal vidéo contenant l'information de luminance à l'instant courant et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ correspondent aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g , x_d , x'_g et x'_d représentent respectivement la position horizontale des sous-régions de fixation centrale et adjacente dans les canaux gauche et droit, et y la position verticale de ces sous-régions dans les deux canaux. Le bloc D se trouve à la figure 5.10, le bloc N à la figure D.12, le bloc U à la figure D.17 et le bloc V à la figure D.18.

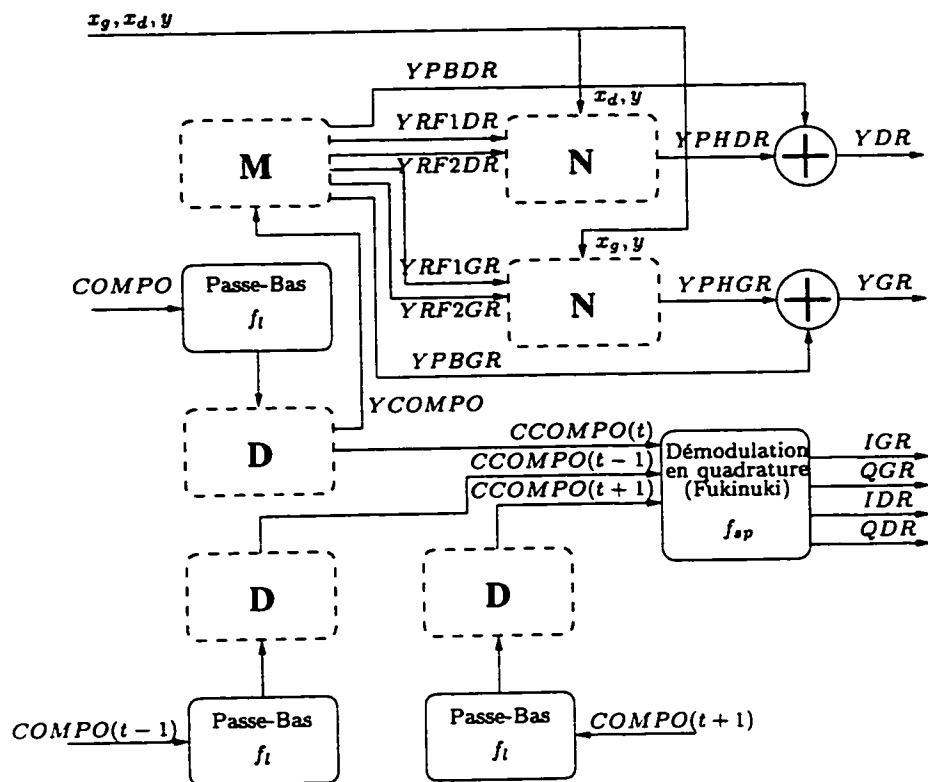


Figure D.16 : Reconstruction des composantes de luminance et de chrominance à partir du signal vidéo composite avec la méthode *intégrée* lorsque la largeur de bande du signal est de 4,2 MHz. $COMPO(t-1)$, $COMPO(t)$ et $COMPO(t+1)$ représentent respectivement le signal vidéo composite à l'instant précédent, l'instant courant et l'instant suivant. De même $CCOMPO(t-1)$, $CCOMPO(t)$ et $CCOMPO(t+1)$ correspondent à la partie du signal vidéo contenant l'information de chrominance à l'instant précédent, l'instant courant et l'instant suivant. $YCOMPO(t)$ représente la partie du signal vidéo contenant l'information de luminance à l'instant courant et $\{YGR, YDR\}$ et $\{IGR, QGR, IDR, QDR\}$ correspondent aux composantes reconstruites de luminance et de chrominance des canaux gauche et droit. x_g , x_d , x'_g et x'_d représentent respectivement la position horizontale des sous-régions de fixation centrale et adjacente dans les canaux gauche et droit, et y la position verticale de ces sous-régions dans les deux canaux. Le bloc D se trouve à la figure 5.10, le bloc M à la figure D.11 et le bloc N à la figure D.12.

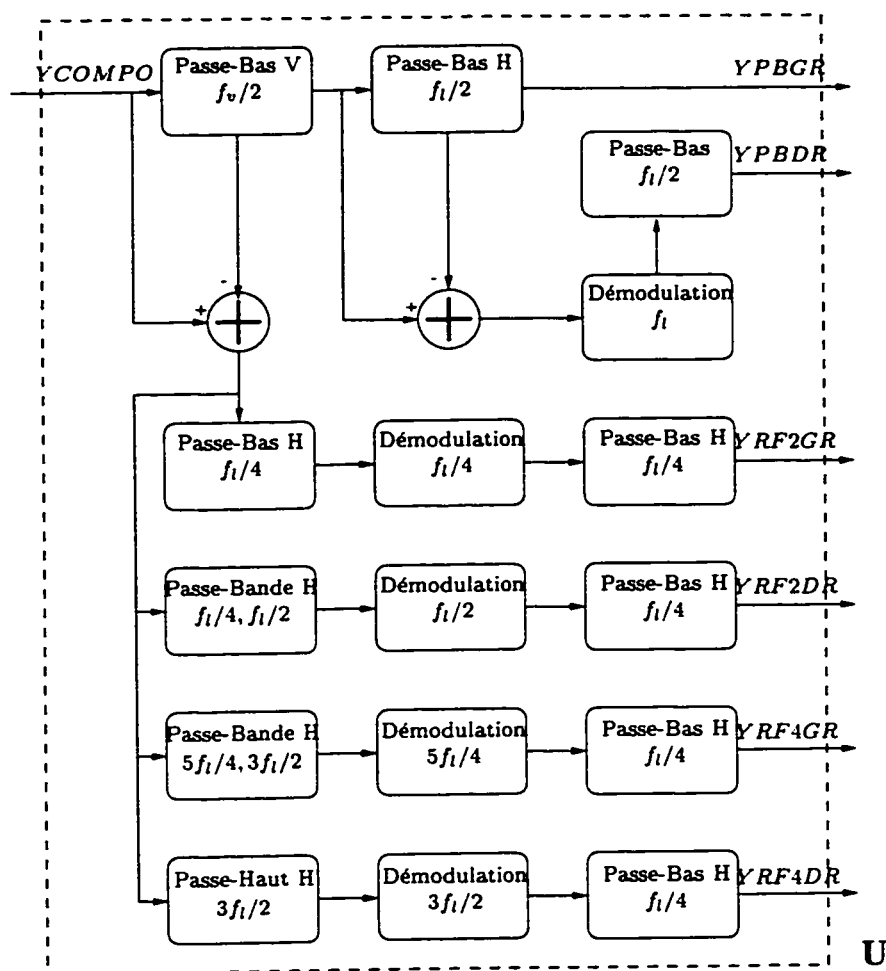


Figure D.17 : Séparation des composantes de luminance avec la méthode *intégrée* (première partie). $YCOMPO$ représente la partie du signal vidéo contenant l'information de luminance et $\{YPBGR, YPBDR\}$, $\{YRF2GR, YRF2DR, YRF4GR, YRF4DR\}$ les composantes de luminance reconstruites passe-bas et les sous-composantes de luminance reconstruites passe-haut des sous-régions de fixation centrale et adjacente pour les canaux gauche et droit.

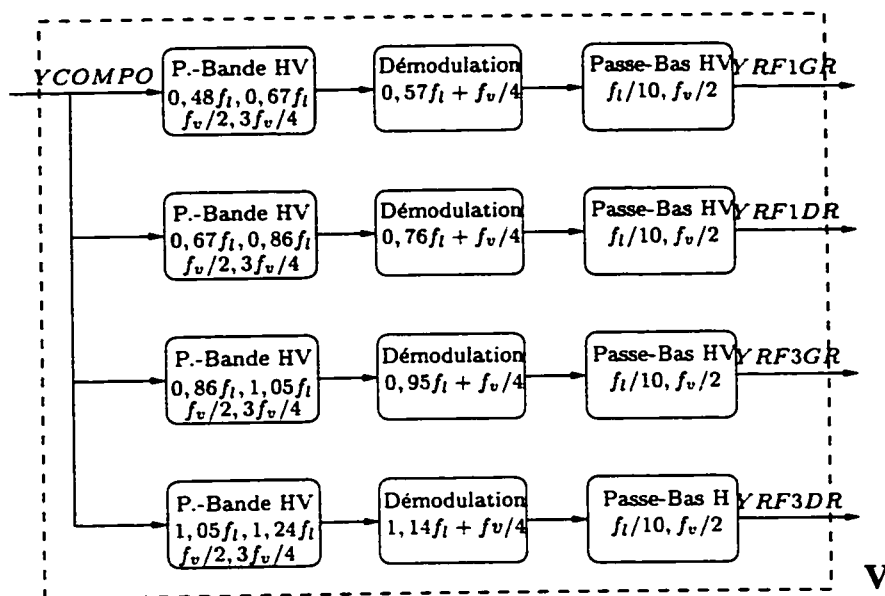


Figure D.18 : Reconstruction des composantes de luminance passe-haut avec la méthode *intégrée* (deuxième partie). $YCOMPO$ représente la partie du signal vidéo contenant l'information de luminance et $\{YPBGR, YPBDR\}$, $\{YRF1GR, YRF1DR, YRF3GR, YRF3DR\}$ les composantes de luminance reconstruites passe-bas et les sous-composantes de luminance reconstruites passe-haut des sous-régions de fixation centrale et adjacente pour les canaux gauche et droit.

à sa position estimée est accompli de façon à tenir compte des résultats de l'étude psychophysique du chapitre 4. Il est donc possible que ce premier décalage donne lieu à un débordement qui, à son tour, causera un second décalage de la région de fixation afin de contrer ce débordement.

Dans le cas de la méthode *intégrée*, les sous-régions de fixation centrale et adjacente doivent toutes les deux être prises en compte lors d'un débordement horizontal. Lorsqu'un tel débordement se produit, la région de fixation n'est pas simplement décalée comme pour les autres méthodes, car cela peut affecter la flexibilité de la représentation qui la rend à la fois adéquate pour une largeur de bande de 4.2 MHz et 7.16 MHz. Par contre, pour un débordement vertical, la même stratégie que pour les méthodes *studio* et *diffusion* est utilisée. Si le débordement est causé par la sous-région adjacente, cette région est décalée, alors que la sous-région centrale ne l'est pas. De cette façon, si le signal vidéo est limité à une largeur de bande de 4.2 MHz, la sous-région centrale n'a pas été modifiée et peut être traitée de la même façon que selon la méthode *diffusion*. Par contre, si le signal vidéo est transmis par un canal avec une bande passante de 7.16 MHz, une partie de la sous-région centrale décalée correspond à une partie de la sous-région centrale et doit donc être éliminée avant la reconstruction de la région de fixation. Si le débordement est causé par la sous-région centrale, alors celle-ci est traitée de la même façon qu'avec la méthode *diffusion*. Dans ce cas, la sous-région adjacente est la même que la sous-région centrale et n'est pas utilisée pour la reconstruction des composantes. La figure D.19 illustre une situation

de débordement horizontal vers la gauche pour les différentes méthodes de codage. Il est clair que le traitement des débordements avec la méthode *intégrée* n'est pas optimal en terme de la taille de la région de fixation transmise, mais c'est le prix à payer pour une représentation flexible qui tient compte des résultats de l'étude des mécanismes de fixation du chapitre 4.

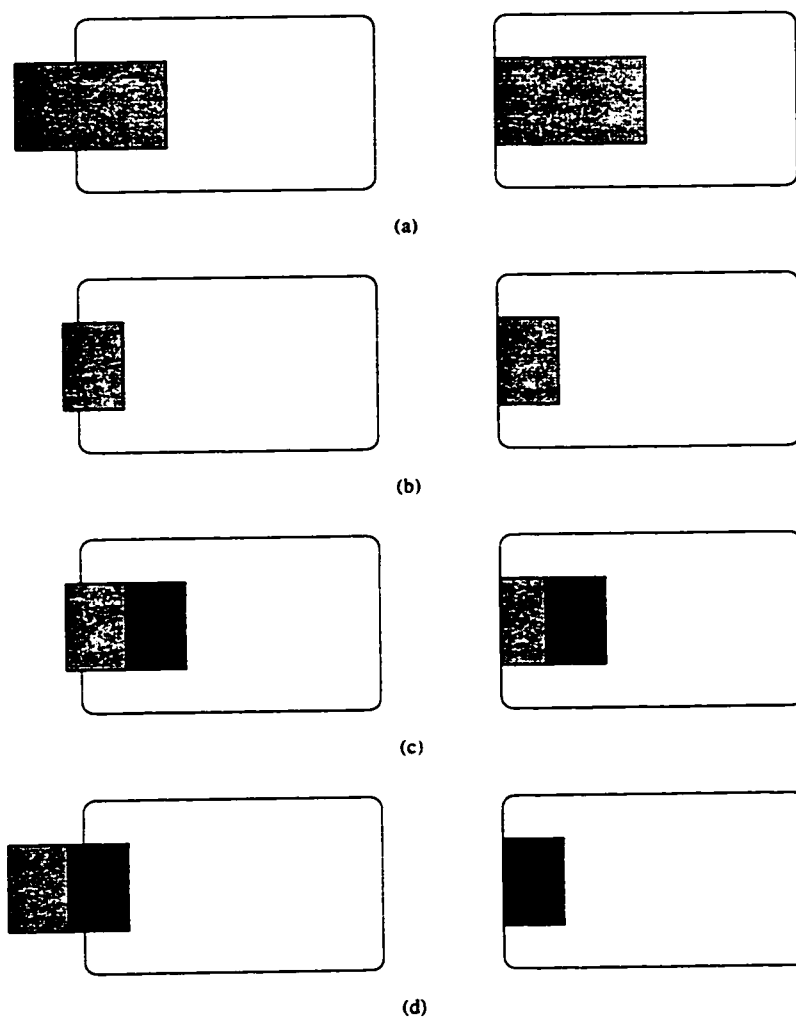


Figure D.19 : Traitement d'un débordement de la région de fixation vers la gauche de l'image : (a) pour le méthode *studio*, (b) pour la méthode *diffusion*, (c) pour la sous-région adjacente de la méthode *intégrée* et (d) pour la sous-région centrale de la méthode *intégrée*.

Annexe E

Données supplémentaires de l'évaluation subjective de la qualité des images codées

E.1 Conditions expérimentales

E.1.1 Système de visualisation

Les séquences codées avec les différentes méthodes du chapitre 5 ont été présentés avec un système de visualisation stéréoscopique qui fonctionne en présentant rapidement à chaque œil les images du point de vue qui lui correspondent à l'aide de lunettes polarisées. Afin d'éviter le papillotements, chaque point de vue est rafraîchi à une fréquence de 60 Hz. Les images présentées doivent être placées selon le format

haut-bas, c'est-à-dire la demi-image gauche dans la partie haute de la mémoire de l'écran et la demi-image droite dans la partie basse de celle-ci. L'écran servant à présenter les images peut être utilisé en mode monoculaire ou stéréoscopique. En mode monoculaire, la mémoire contenant les deux images est balayée à une fréquence de 60 Hz, avec l'ordre de balayage des lignes en mémoire correspondant à celui de l'affichage des lignes à l'écran. Avec le format *haut-bas* en mode monoculaire, deux images superposées sont donc visibles à l'écran. Lorsque l'écran est mis en mode stéréo, la mémoire est toujours balayée à 60 Hz, mais la correspondance entre les lignes mémoire et les lignes écran est maintenant changée de façon à présenter une demi-image à la taille du plein écran en affichant seulement une ligne sur deux. Donc, en $\frac{1}{120}$ de seconde, une demi-image est balayée en mémoire et est affichée à l'écran. Durant le $\frac{1}{120}$ de seconde suivant, la seconde demi-image est balayée en mémoire et est affichée à l'écran, sur les mêmes lignes que l'image précédente. Donc, les images sont en fait montrées à une fréquence de 120 Hz, avec les points de vue gauche et droit présentés en alternance chacun à une fréquence de 60 Hz. Les lunettes polarisées sont synchronisées avec l'écran de façon à ce que le point de vue droit ne soit visible que de l'œil droit et le point de vue gauche que de l'œil gauche. La figure E.1 illustre le fonctionnement du système de visualisation stéréoscopique.

Il est important de réaliser qu'il existe des systèmes commerciaux stéréoscopiques basés sur le format NTSC qui transmettent l'information du point de vue gauche dans les champs pairs et celle du point de vue droit dans les champs impairs. Na-

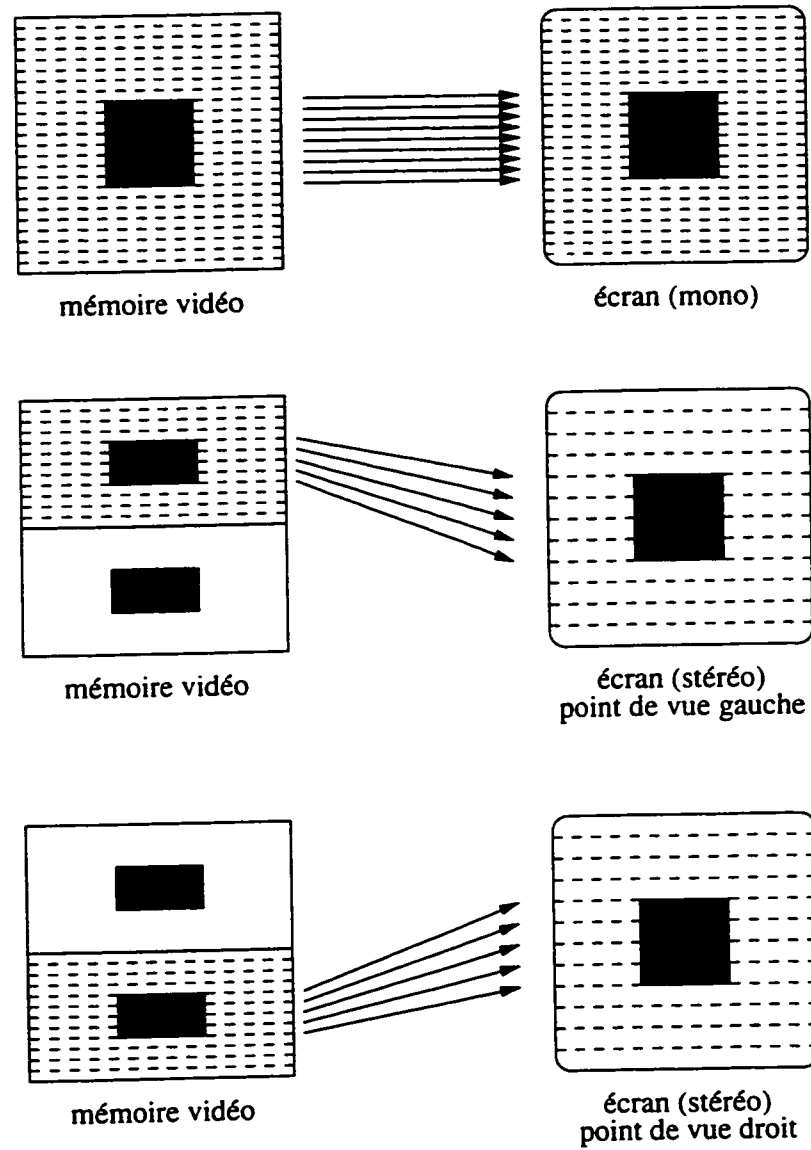


Figure E.1 : Principe de fonctionnement du système de visualisation stéréoscopique.

turellement, la résolution verticale est réduite par deux et chaque point de vue est présenté à 30 Hz. Une des justifications du développement des nouvelles méthodes de codage stéréoscopiques du chapitre 5 provient de la transmission de l'information stéréoscopique complète à chaque champ, une nette amélioration par rapport aux systèmes commerciaux. De plus, ces nouvelles méthodes peuvent aussi être utilisées avec d'autres types de systèmes de visualisation, comme des casques de réalité virtuelle ou des systèmes avec deux écrans pleine résolution combinés à un système optique limitant le point de vue de chaque œil à un des deux écrans.

Par contre, avec le présent système de visualisation, même si des images complètes sont disponibles, des demi-images doivent être utilisées. Ceci correspond donc à utiliser un champ sur deux, de la même façon que les systèmes commerciaux, mais tout de même avec une fréquence temporelle deux fois plus élevée que ces systèmes. Donc, même si les méthodes de codage présentées au chapitre 5 permettent la transmission d'une plus grande quantité d'information que les systèmes commerciaux, les limites du système de visualisation quant à la résolution verticale des images ne nous permet pas d'exploiter pleinement tous les avantages de ces nouvelles méthodes de codage.

E.1.2 Installation

La taille de l'écran sur lequel les séquences ont été présentées est de 19 pouces. Puisque la distance de l'écran à laquelle la fusion stéréoscopique est accomplie varie selon les sujets, une position de visualisation fixe ne leur a pas été imposée. Il a

plutôt été demandé aux sujets, lors d'une visualisation préliminaire, d'établir une position par rapport à l'écran à laquelle ils se sentaient à l'aise pour fusionner les images stéréoscopiques. Par la suite, chaque sujet a conservé cette position pour toute la période de test. La distance entre l'écran et les sujets s'étendait approximativement de 30 cm à 80 cm selon les sujets. L'étape préliminaire de visualisation permettait de s'assurer que les sujets avaient une fusion stéréoscopique normale et pouvaient donc accomplir le test. L'éclairage de la pièce était d'environ 40 cd m^{-2} . Les sujets consignaient leur réponse sur une feuille de papier après la visualisation de chaque séquence, en notant leur appréciation en fonction de l'échelle de visibilité de dégradations du tableau 6.1 et en donnant une description qualitative des dégradations observées.

E.1.3 Stimuli

Quarante séquences stéréoscopiques ont été présentées à chaque sujet, soit les cinq séquences de la figure 6.1 codées chacune avec les six méthodes du chapitre 5 et les cinq séquences originales non-traitées présentées chacune à deux reprises. La présentation se faisait par séquence, avec l'ordre de présentation des séquences choisi de façon aléatoire pour chaque sujet. Pour chaque séquence, l'ordre de présentation des 8 conditions (6 codées et 2 originales) était aussi choisi de façon aléatoire. Au début de chaque séquence, un attracteur d'attention était présenté. Il consistait en un point blanc sur un fond noir, positionné au centre de la région de fixation. Ce point

était visible, s'éteignait brièvement et réapparaissait à la même position avant le début de la présentation des images de la séquence. Toutes les séquences étaient présentées de façon continue, c'est-à-dire qu'après la dernière image, l'attracteur d'attention réapparaissait, clignotait de nouveau et les images de la séquence défilaient par la suite.

E.1.4 Méthode

Trois sujets avec une vision stéréoscopique normale ont été testés. Pour chaque bloc de 8 séquences, la séquence originale non-traitée était tout d'abord présentée. Les sujets étaient informés que cette séquence devait être utilisée comme la référence par rapport à laquelle le niveau de qualité des autres séquences serait évalué. Pour chaque séquence, la région de fixation était clairement décrite aux sujets. Pour les séquences où la position de la région de fixation demeurait fixe, il était demandé aux sujets de fixer cette région. Pour celle où la région de fixation se déplaçait, il était plutôt demandé aux sujets de fixer un objet d'intérêt situé à la position de l'attracteur d'attention et à suivre les déplacements de cet objet au cours de la séquence. Puisque les séquences étaient présentées en boucle, les sujets avaient la possibilité de les observer aussi longtemps que nécessaire. Après la visualisation de chaque séquence, les sujets consignaient par écrit leur appréciation du niveau de qualité de la séquence. Les sujets avaient été informés que 8 méthodes de codage étaient testées et ignoraient donc que certaines conditions correspondaient aux séquences originales. Les sujets

avaient la possibilité de revoir sur demande les séquences originales de façon à *re-calibrer* leur référence de comparaison si nécessaire. La durée approximative de la procédure de test était de deux heures. La longueur de la procédure de test provient de la durée de chargement en mémoire de chaque séquence, qui varie de une à quatre minutes selon la longueur de la séquence. De plus, avant l'affichage de la séquence à pleine vitesse, le programme de visualisation affiche les images de la séquence une à la suite de l'autre à une cadence réduite. Afin de ne pas biaiser leur jugement, les sujets ne regardaient pas l'écran durant cet affichage.

Il est clair que, tout comme l'étude du chapitre 4, cette étude sur le niveau de qualité des images codée ne prétend pas être une étude psychophysique d'envergure permettant d'évaluer systématiquement la cause de chaque type de dégradation introduite par les méthodes de codage. Elle doit plutôt être considérée comme une étude informelle, quoi que tout à fait rigoureuse quant à la procédure expérimentale choisie, qui permet de fournir rapidement l'information nécessaire à une évaluation de la performance globale méthodes de codage du chapitre 5. Une évaluation psychophysique plus poussée impliquerait un plus grand nombre de sujets, la présentation des différentes conditions de codage un plus grand nombre de fois pour chaque séquence (et en conséquence une procédure de test beaucoup plus longue) et un suivi du mouvement des yeux des sujets afin de s'assurer que leur point de fixation demeure dans la région de fixation.

E.2 Résultats

L'appréciation des trois sujets en fonction de l'échelle du tableau 6.1 est présentée aux tableaux E.1, E.2 et E.3 et l'appréciation moyenne de ces trois sujets se trouve au tableau 6.2. Lors de leur évaluation, les sujets avaient quelquefois de la difficulté à choisir entre deux niveaux de l'échelle. En conséquence, les sujets ont souvent attribué des *demi-notes* qui sont représentées dans les tableaux E.1, E.2 et E.3 à l'aide du signe +. Ainsi, par exemple, une note de 2+ représente une image dont le niveau de qualité est jugé moins bon qu'une image de niveau 1, mais meilleur qu'une image de niveau 2, et sa pondération pour le calcul des appréciations moyennes est donc de 1,5.

Les sujets CZ et AB ont noté la qualité des séquences principalement en fonction du ternissement des couleurs et de l'augmentation du flou des images, alors que le sujet JYH était beaucoup plus sensible aux détails spatiaux des objets des différentes séquences. Ainsi, dans son évaluation il a accordé plus d'importance à un défaut se traduisant par le papillotement de lignes horizontales ou verticales à la frontière de la région de fixation, principalement pour les séquences *piano* et *tunnel*. Ce défaut n'a pu être corrigé avant de tester les sujets, car il ne se manifeste que lors d'une présentation des séquences codées à une vitesse permettant la perception de mouvements continus dans les séquences, et que la visualisation de ces séquences à cette vitesse à l'aide de notre système n'a été possible que seulement quelques jours avant le test. Ce défaut provient probablement des conditions frontières utilisées avec les filtres, puisqu'il se manifeste aussi de temps en temps dans des régions peu texturées

Tableau E.1 : Appréciation du niveau de qualité des images pour le sujet CZ.

	or.	st.	st. F.	dif.	dif. F.	int. 4,2	int. 7,16
aqua	1/1	2+	1	2	2	2	3
manège	2/1	2	2	4	4	4	4
piano	1/1	1	3	4	2	3	2
train	1/1	1	1	2+	2+	3	3
tunnel	1/1	2+	2+	4	3	3	3
moyenne	1,1	1,4	1,7	3,1	2,5	3,0	3,0

situées aux frontières des images, bien que, lorsqu'il se trouve dans ces régions, il est beaucoup moins remarqué par les sujets. La raison pour laquelle ce défaut n'est observé que pour les séquences *piano* et *tunnel* provient du fait que ce sont les deux séquences où la position de la région de fixation est fixe. Une solution simple qui permettrait probablement d'éliminer ce défaut consiste, plutôt qu'à utiliser toujours la même position pour la région de fixation lorsque celle-ci est fixe, à légèrement perturber de façon aléatoire sa position à chaque champ. Étant donnée la grande taille de la région de fixation, ceci ne devrait pas affecter significativement la perception des images tout en rendant beaucoup moins visible la démarcation entre la région de fixation et le fond.

Les méthodes *studio* et *studio* avec trous de Fukinuki reçoivent en général la meilleure appréciation. La seule exception provient de la séquence *tunnel* évaluée par le sujet JYH qui a été notée assez faiblement, justement en raison de la présence de papillotements à la démarcation de la région de fixation. Cette exception fait significativement augmenter la moyenne du sujet JYH pour les conditions *studio* et

Tableau E.2 : Appréciation du niveau de qualité des images pour le sujet JYH.

	or.	st.	st. F.	dif.	dif. F.	int. 4,2	int. 7,16
aqua	1/1	2	2	3+	4	3	3
manège	1/1	3+	4	4+	3	3+	2
piano	1/1	2	4	5+	3+	3+	3+
train	1/1	3+	3	3+	2+	5+	3
tunnel	2/1	6+	5	4	4+	3+	3
moyenne	1,1	2,9	3,6	3,4	2,9	3,0	2,7

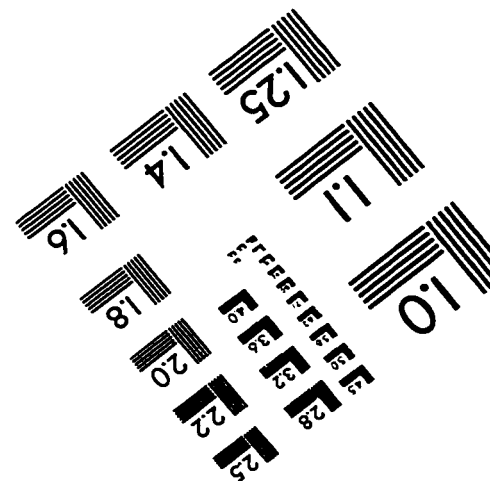
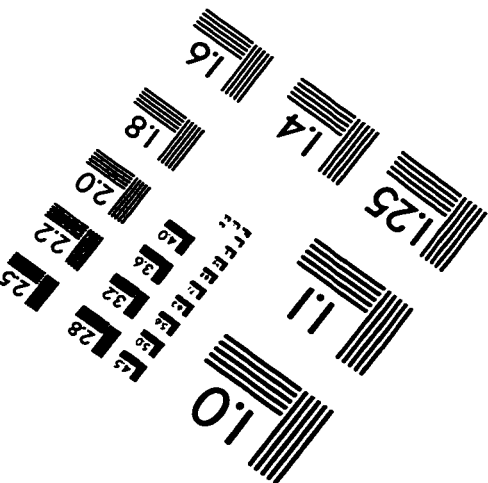
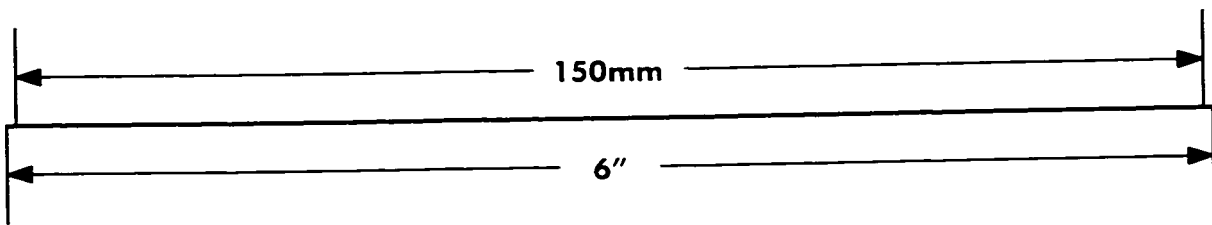
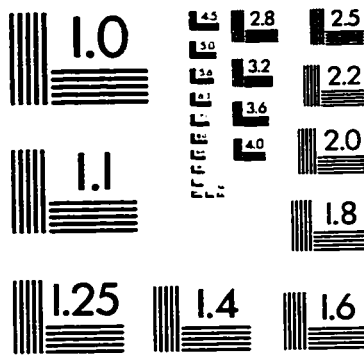
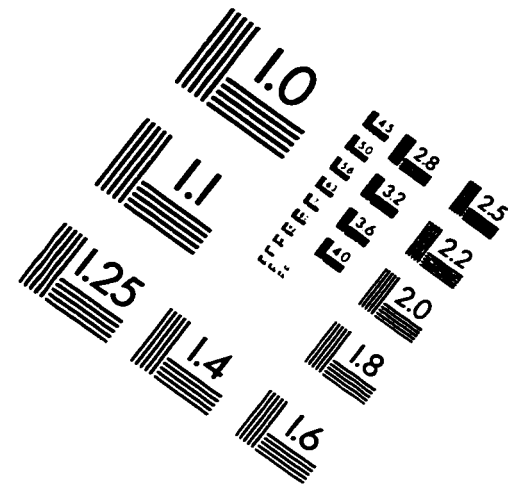
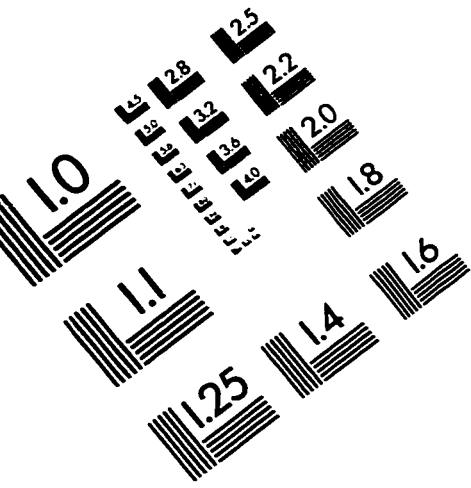
Tableau E.3 : Appréciation du niveau de qualité des images pour le sujet AB.

	or.	st.	st. F.	dif.	dif. F.	int. 4,2	int. 7,16
aqua	2/1	1	1	2	3	1	3
manège	2/1	3	2	4	4	3	2
piano	1/1	1	2	2	3	1	1
train	1/1	2	1	1	3	4	3
tunnel	1/1	2	1	2	4	2	3
moyenne	1,3	1,8	1,4	2,2	3,4	2,2	2,4

studio avec trous de Fukinuki, bien que les autres séquences codées avec ces conditions reçoivent une bonne appréciation. L'évaluation des méthodes *diffusion*, *diffusion* avec trous de Fukinuki, *intégrée* limitée à 4.2 MHz et *intégrée* limitée à 7.16 MHz varie selon les sujets. En moyenne, ces méthodes peuvent être considérées équivalentes en tenant compte de la variance statistique des données.

Les sujets JYH et CZ ont aussi identifié des légers dédoublements de contours à l'intérieur de la région de fixation. Ce défaut est inhérent à l'approche de codage choisie qui implique l'addition de composantes de luminance passe-haut et passe-bas. En raison de la non-idéalité des filtres, certaines distorsions sont introduites lors du filtrage des composantes. Ces distorsions peuvent faire que, lors de l'addition des composantes, certaines entités de l'image ne sont plus tout à fait alignées, ce qui se traduit par un dédoublement de contours. Cet effet dépend des filtres qui sont utilisés ainsi que du nombre d'opérations de filtrage accomplies pour chaque méthode de codage.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved